

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

seq_documentation_block:

```
; Sequence 29, Application US/08649991
; Patent No. 5919462
; GENERAL INFORMATION:
; APPLICANT: Narwa, Remy
; APPLICANT: Roques, Pierre
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
; STREET: 1800 M Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-649-991-29

alignment_scores:
    Quality: 9.00      Length: 9
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-08-649-991-29/rev ..
Align seg 1/1 to reverse of: US-08-649-991-29 from: 1 to: 54

972 LeuphSerThrSerArgAlaSer 980
|||||
29 CTCCTTTCCACTTCCTCTAGAGCTTCT 3

seq_name: /cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-130

seq_documentation_block:
; Sequence 130, Application US/08649991
; Patent No. 5919462
; GENERAL INFORMATION:
; APPLICANT: Narwa, Remy
; APPLICANT: Roques, Pierre
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130

; Sequence 29, Application US/08649991
; Patent No. 5919462
; GENERAL INFORMATION:
; APPLICANT: Narwa, Remy
; APPLICANT: Roques, Pierre
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
; STREET: 1800 M Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ ID NO: 130:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-649-991-130

alignment_scores:
    Quality: 9.00      Length: 9
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-08-649-991-130/rev ..
Align seg 1/1 to reverse of: US-08-649-991-130 from: 1 to: 54

972 LeuphSerThrSerArgAlaSer 980
|||||
29 CTCCTTTCCACTTCCTCTAGAGCTTCT 3

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5219739-8

seq_documentation_block:
; Patent No. 5219739
; APPLICANT: FISCHER, EDMUND G.; ABRAHAM, JUDITH A.; FIDDES,
; JOHN C.; MITCHELL, RICHARD L.
; TITLE OF INVENTION: DNA SEQUENCES ENCODING BVGEF120 AND
; BVGEF 121 AND METHODS FOR THE PRODUCTION OF BOVINE AND HUMAN
; VAASCULAR ENDOTHELIAL CELL GROWTH FACTORS, BVGEF120 AND BVGEF121
; NUMBER OF SEQUENCES: 40
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/559,041
; FILING DATE: 27-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 450,883
; FILING DATE: 14-DEC-1989
; APPLICATION NUMBER: 387,545
; FILING DATE: 27-JUL-1989
; SEQ ID NO: 8
; LENGTH: 789
; 5219739-8
```



```
alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x 5219739-8 ..
Align seg 1/1 to: 5219739-8 from: 1 to: 789

227 ProProHishHishHishHishHish 235
|||||
416 CCACCACACACACACACACACCAT 442

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5194596-16

seq_documentation_block:
; Patent No. 5194596
; APPLICANT: TISCHER, EDMUND G.; ABRAHAM, JUDITH A.; FIDDES, JOHN
; C.: MITCHELL, RICHARD L.
; TITLE OF INVENTION: PRODUCTION OF VASCULAR ENDOTHELIAL CELL
; GROWTH FACTOR
; NUMBER OF SEQUENCES: 32
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/450,883
; FILING DATE: 14-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 387,545
; FILING DATE: 27-JUL-1989
; SEQ ID NO: 16:
; LENGTH: 961
5194596-16

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x 5194596-16 ..
Align seg 1/1 to: 5194596-16 from: 1 to: 961

227 ProProHishHishHishHishHish 235
|||||
588 CCACCACACACACACACACACCAT 614

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5219739-16

seq_documentation_block:
; Patent No. 5219739
; APPLICANT: TISCHER, EDMUND G.; ABRAHAM, JUDITH A.; FIDDES,
; JOHN C.; MITCHELL, RICHARD L.
; TITLE OF INVENTION: DNA SEQUENCES ENCODING BVGEF120 AND
; BVGEF 121 AND METHODS FOR THE PRODUCTION OF BOVINE AND HUMAN
; VASCULAR ENDOTHELIAL CELL GROWTH FACTORS, BVGEF120 AND BVGEF121
; NUMBER OF SEQUENCES: 40
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/559,041
; FILING DATE: 27-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 450,883
; FILING DATE: 14-DEC-1989
; APPLICATION NUMBER: 387,545
; FILING DATE: 27-JUL-1989
; SEQ ID NO: 16:
; LENGTH: 961
5219739-16

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x 5219739-16 from: 1 to: 961

227 ProProHishHishHishHishHish 235
|||||
588 CCACCACACACACACACACACCAT 614

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-211-942-16

seq_documentation_block:
; Sequence 16, Application US/08211942
; Patent No. 523287
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Thomas
; APPLICANT: Bialojan, Siegfried
; APPLICANT: Kroeger, Burkhard
; APPLICANT: Kuenast, Christoph
; TITLE OF INVENTION: No. 523287el thrombin-inhibitory protein from assassin
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinlauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/211,942
; FILING DATE:
; CLASSIFICATION: 435
; CLASSIFICATION: C07K 13/00
; CLASSIFICATION: A61K 37/64
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP92/02450
; FILING DATE: 27-OCT-1992
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-211-942-16

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-08-211-942-16 ..
Align seg 1/1 to: US-08-211-942-16 from: 1 to: 1282

227 ProProHishHishHishHishHish 235
|||||
549 CCTCCTCCTCATCATCATCATCAT 575

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5332671-5

seq_documentation_block:
; Patent No. 5332671
```

```
Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x 5219739-16 ..
Align seg 1/1 to: 5219739-16 from: 1 to: 961

227 ProProHishHishHishHishHish 235
|||||
588 CCACCACACACACACACACACCAT 614

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-211-942-16

seq_documentation_block:
; Sequence 16, Application US/08211942
; Patent No. 523287
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Thomas
; APPLICANT: Bialojan, Siegfried
; APPLICANT: Kroeger, Burkhard
; APPLICANT: Kuenast, Christoph
; TITLE OF INVENTION: No. 523287el thrombin-inhibitory protein from assassin
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinlauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/211,942
; FILING DATE:
; CLASSIFICATION: 435
; CLASSIFICATION: C07K 13/00
; CLASSIFICATION: A61K 37/64
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP92/02450
; FILING DATE: 27-OCT-1992
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-211-942-16

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-08-211-942-16 ..
Align seg 1/1 to: US-08-211-942-16 from: 1 to: 1282

227 ProProHishHishHishHishHish 235
|||||
549 CCTCCTCCTCATCATCATCATCAT 575

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5332671-5

seq_documentation_block:
; Patent No. 5332671
```

```
; APPLICANT: FERRARA, NAPOLEONE;LEUNG, DAVID W.H.
; TITLE OF INVENTION: PRODUCTION OF VASCULAR ENDOTHELIAL CELL
; GROWTH FACTOR AND DNA ENCODING SAME
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/389,722
; FILING DATE: 04-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 369,424
; FILING DATE: 21-JUN-1989
; APPLICATION NUMBER: 351,117
; FILING DATE: 12-MAY-1989
; SEQ ID NO:5:
; LENGTH: 1543
5332671-5

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x 5332671-5 ..
Align seg 1/1 to: 5332671-5 from: 1 to: 1543

227 ProProHSHSHSHSHSHSHSHS 235
|||||
1211 CCACCACCACACACCACCACCACCAT 1237

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-879-647A-25

seq_documentation_block:
; Sequence 25, Application US/07879647A
; Patent No. 5266689
; GENERAL INFORMATION:
; APPLICANT: Chakraborty, P.R.
; APPLICANT: Dashkevich, M.
; APPLICANT: Elbrecht, A.
; APPLICANT: Feighner, S.D.
; APPLICANT: Liberator, P.A.
; APPLICANT: Profous-Juchelka, H.
; TITLE OF INVENTION: Eimeria Maxima DNA
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
; MEDIUM TYPE: storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.4
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/879,647A
; FILING DATE: 19920512
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/706,628
; FILING DATE: 29-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tribble, Jack L.
; REGISTRATION NUMBER: 32,633
; REFERENCE/DOCKET NUMBER: .184201A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-5321
; TELEFAX: (908) 594-4720
```

```
; TELEX: 138825
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1744 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-879-647A-25

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-07-879-647A-25 ..
Align seg 1/1 to: US-07-879-647A-25 from: 1 to: 1744

258 ThrLysProThrHisLeuValgluPro 266
|||||
193 ACAAAACCAACCCACCTGTGGAGCCT 219

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-879-584A-25

seq_documentation_block:
; Sequence 25, Application US/07879584A
; Patent No. 5278298
; GENERAL INFORMATION:
; APPLICANT: Chakraborty, P.R.
; APPLICANT: Dashkevich, M.
; APPLICANT: Elbrecht, A.
; APPLICANT: Feighner, S.D.
; APPLICANT: Liberator, P.A.
; APPLICANT: Profous-Juchelka, H.
; TITLE OF INVENTION: Eimeria Brunetti DNA
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
; MEDIUM TYPE: storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.4
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/879,584A
; FILING DATE: 19920512
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/706,717
; FILING DATE: 29-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tribble, Jack L.
; REGISTRATION NUMBER: 32,633
; REFERENCE/DOCKET NUMBER: .184191A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-5321
; TELEFAX: (908) 594-4720
; TELEX: 138825
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1744 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
```

US-07-879-584A-25

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x US-07-879-584A-25 ..

Align seg 1/1 to: US-07-879-584A-25 from: 1 to: 1744

258 ThrLysProThrHisLeuValGluPro 266

|||||
193 ACAAAACCAACCCACCTGTGTGGAGCCT 219

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-879-470A-25

seq_documentation_block:

: Sequence 25, Application US/07879470A
: Patent No. 5288845

: GENERAL INFORMATION:

: APPLICANT: Chakraborty, P.R.

: APPLICANT: Dashkevich, M.

: APPLICANT: Elbrecht, A.

: APPLICANT: Feighner, S.D.

: APPLICANT: Liberator, P.A.

: APPLICANT: Profous-Juchelka, H.

: TITLE OF INVENTION: Eimeria Necatrix DNA

: TITLE OF INVENTION: Probes

: NUMBER OF SEQUENCES: 50

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Merck & Co., Inc.

: STREET: 126 Lincoln Avenue

: CITY: Rahway

: STATE: New Jersey

: COUNTRY: USA

: ZIP: 07065

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb

: MEDIUM TYPE: storage

: COMPUTER: Apple Macintosh

: OPERATING SYSTEM: Macintosh 6.0.4

: SOFTWARE: Microsoft Word 4.0

: CURRENT APPLICATION DATA: US/07/879,470A

: APPLICATION NUMBER: 07/706,351

: FILING DATE: 19920512

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: 07/706,351

: FILING DATE: 29-MAY-1991

: ATTORNEY/AGENT INFORMATION:

: NAME: Tribble, Jack L.

: REGISTRATION NUMBER: 32,633

: REFERENCE/DOCKET NUMBER: 184221A

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (908) 594-5321

: TELEFAX: (908) 594-4720

: TELEX: 138825

: INFORMATION FOR SEQ ID NO: 25:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 1744 bases

: TYPE: NUCLEIC ACID

: STRANDEDNESS: single

: TOPOLOGY: linear

US-07-879-470A-25

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x US-07-879-470A-25 ..

Align seg 1/1 to: US-07-879-470A-25 from: 1 to: 1744

258 ThrLysProThrHisLeuValGluPro 266

|||||
193 ACAAAACCAACCCACCTGTGTGGAGCCT 219

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-879-644A-25

seq_documentation_block:

: Sequence 25, Application US/07879644A

: Patent No. 5298613

: GENERAL INFORMATION:

: APPLICANT: Chakraborty, P.R.

: APPLICANT: Dashkevich, M.

: APPLICANT: Elbrecht, A.

: APPLICANT: Feighner, S.D.

: APPLICANT: Liberator, P.A.

: APPLICANT: Profous-Juchelka, H.

: TITLE OF INVENTION: Eimeria Acervulina DNA

: TITLE OF INVENTION: Probes

: NUMBER OF SEQUENCES: 50

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Merck & Co., Inc.

: STREET: 126 Lincoln Avenue

: CITY: Rahway

: STATE: New Jersey

: COUNTRY: USA

: ZIP: 07065

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb

: MEDIUM TYPE: storage

: COMPUTER: Apple Macintosh

: OPERATING SYSTEM: Macintosh 6.0.4

: SOFTWARE: Microsoft Word 4.0

: CURRENT APPLICATION DATA: US/07/879,644A

: APPLICATION NUMBER: 07/706,817

: FILING DATE: 19920512

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: 07/706,817

: FILING DATE: 29-MAY-1991

: ATTORNEY/AGENT INFORMATION:

: NAME: Tribble, Jack L.

: REGISTRATION NUMBER: 32,633

: REFERENCE/DOCKET NUMBER: 184181A

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (908) 594-5321

: TELEFAX: (908) 594-4720

: TELEX: 138825

: INFORMATION FOR SEQ ID NO: 25:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 1744 bases

: TYPE: NUCLEIC ACID

: STRANDEDNESS: single

: TOPOLOGY: linear

US-07-879-644A-25

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x US-07-879-644A-25 ..

Align seg 1/1 to: US-07-879-644A-25 from: 1 to: 1744

258 ThrLysProThrHisLeuValGluPro 266

```
|||||
193 ACAAACCAACCCACCTGTGGAGCCT 219

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-879-640A-25
seq_documentation_block:
; Sequence 25, Application US/07879640A
; Patent No. 5359050
; GENERAL INFORMATION:
; APPLICANT: Chakraborty, P.R.
; APPLICANT: Dashkevicz, M.
; APPLICANT: Elbrecht, A.
; APPLICANT: Feighner, S.D.
; APPLICANT: Liberator, P.A.
; APPLICANT: Profous-Juchelka, H.
; TITLE OF INVENTION: Eimeria Mitis DNA
; TITLE OF INVENTION: Probes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
; MEDIUM TYPE: storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.4
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/879,640A
; FILING DATE: 19920512
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/706,355
; FILING DATE: 29-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tribble, Jack L.
; REGISTRATION NUMBER: 32,633
; REFERENCE/DOCKET NUMBER: .18421IA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-5321
; TELEFAX: (908) 594-4720
; TELEX: 138825
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1744 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-879-640A-25

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-07-879-640A-25 ..
Align seg 1/1 to: US-07-879-640A-25 from: 1 to: 1744

258 ThrLysProThrHisLeuValGluPro 266
|||||
193 ACAAACCAACCCACCTGTGGAGCCT 219

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-879-594A-25
seq_documentation_block:
; Sequence 25, Application US/07879594A
```

```
; Patent No. 5449768
; GENERAL INFORMATION:
; APPLICANT: Chakraborty, P.R.
; APPLICANT: Dashkevicz, M.
; APPLICANT: Elbrecht, A.
; APPLICANT: Feighner, S.D.
; APPLICANT: Liberator, P.A.
; APPLICANT: Profous-Juchelka, H.
; TITLE OF INVENTION: Eimeria Praecox DNA
; TITLE OF INVENTION: Probes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
; MEDIUM TYPE: storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.4
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/879,594A
; FILING DATE: 19920512
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/706,360
; FILING DATE: 29-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tribble, Jack L.
; REGISTRATION NUMBER: 32,633
; REFERENCE/DOCKET NUMBER: .18423IA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-5321
; TELEFAX: (908) 594-4720
; TELEX: 138825
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1744 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-879-594A-25

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-07-879-594A-25 ..
Align seg 1/1 to: US-07-879-594A-25 from: 1 to: 1744

258 ThrLysProThrHisLeuValGluPro 266
|||||
193 ACAAACCAACCCACCTGTGGAGCCT 219

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-879-469A-25
seq_documentation_block:
; Sequence 25, Application US/07879469A
; Patent No. 5563256
; GENERAL INFORMATION:
; APPLICANT: Chakraborty, P.R.
; APPLICANT: Dashkevicz, M.
; APPLICANT: Elbrecht, A.
; APPLICANT: Feighner, S.D.
; APPLICANT: Liberator, P.A.
```

```

; APPLICANT: Profous-Juchelka, H.
; TITLE OF INVENTION: Eimeria Tenella DNA
; TITLE OF INVENTION: Probes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
; MEDIUM TYPE: storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.4
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/879,469A
; FILING DATE: 19920512
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/706,362
; FILING DATE: 29-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tribble, Jack L.
; REGISTRATION NUMBER: 32,633
; REFERENCE/DOCKET NUMBER: 184241A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-5321
; TELEFAX: (908) 594-4720
; TELEX: 138825
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1744 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-879-469A-25

```

```

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```
alignment_block:
US-09-332-522B-2 x US-07-879-469A-25 ..

```

```
Align seg 1/1 to: US-07-879-469A-25 from: 1 to: 1744

```

```

258 ThrLysProThrHisLeuValGluPro 266
|||||
193 ACAAAACCAACCCACCTGTGGAGCCT 219

```

```
seq_name: /cgn2_6/ptodata/2/1na/PTCUTS_COMB.seq:PCT-US92-09202-3

```

```

; seq_documentation_block:
; Sequence 3, Application PC/TUS9209202
; GENERAL INFORMATION:
; APPLICANT: Hogan, Michael E.
; TITLE OF INVENTION: Triplex Forming Oligonucleotide Reagents Targeted
; TITLE OF INVENTION: to the Neu Oncogene Promoter and Method of Use
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski
; STREET: 1301 McManney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/09202
; FILING DATE: 19921028
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5393
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5151
; TELEFAX: 713/651-5246
; TELEX: 762829
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Triplex forming oligonucleotide
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
PCT-US92-09202-3

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x PCT-US92-09202-3/rev ..
Align seg 1/1 to reverse of: PCT-US92-09202-3 from: 1 to: 41

228 ProProHisHisHisHisHis 235
|||||
32 CCACCCACCCACCCACCCACCAC 9

seq_name: /cgn2_5/ptodata/2/1na/5D_COMB.seq:US-08-911-774-1
seq_documentation_block:
; Sequence 1, Application US/08911774
; Patent No. 5968781
; GENERAL INFORMATION:
; APPLICANT: YOON, Sung-June
; APPLICANT: KIM, Jong-Woo
; APPLICANT: HUM, Yong
; APPLICANT: RHO, Hyune-Mo
; APPLICANT: JUNG, Gu-Hung
; TITLE OF INVENTION: HBV Polymerase, RNase H Enzyme Derived
; TITLE OF INVENTION: From HBV Polymerase, Processes For Preparation And Uses
; TITLE OF INVENTION: For Screening Antiviral Agents Thereof
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BLVD.
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: USA
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,774
; FILING DATE: 15-AUG-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: STEEN, JEFFREY S
; REGISTRATION NUMBER: 32063
; REFERENCE/DOCKET NUMBER: 685-13 (7FPO-07-03)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-228-8484
; TELEFAX: 516-228-8516
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA (synthetic
; DESCRIPTION: oligonucleotide)"
US-08-911-774-1

alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-08-911-774-1 ..
Align seg 1/1 to: US-08-911-774-1 from: 1 to: 50

228 ProProHishHishHishHis 235
|||||
6 CCACCGCATCACCACCATCACC 29

seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-228-243-1

seq_documentation_block:
; Sequence 1, Application US/09228243
; Patent No. 6071734
; GENERAL INFORMATION:
; APPLICANT: Yoon, Sung-June
; APPLICANT: Kim, Jong-woo
; APPLICANT: Ruh, Yong
; APPLICANT: Rho, Hyune-Mo
; APPLICANT: Jung, Gu-Hung
; TITLE OF INVENTION: HBV Polymerase, RNase H enzyme derived
; TITLE OF INVENTION: from HBV polymerase, processes for preparation and uses
; TITLE OF INVENTION: for screening antiviral agents thereof
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/228,243
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Steen, Jeffrey S.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic oligonucleotide"
US-09-228-243-1

alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-09-228-243-1 ..
Align seg 1/1 to: US-09-228-243-1 from: 1 to: 50

228 ProProHishHishHishHis 235
|||||
6 CCACCGCATCACCACCATCACC 29

seq_name: /cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-30

seq_documentation_block:
; Sequence 30, Application US/08649991
; Patent No. 5919462
; GENERAL INFORMATION:
; APPLICANT: Narwa, Remy
; APPLICANT: Roques, Pierre
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; TITLE OF INVENTION: HIV-1 VIRUS GENOME. CORRESPONDING PEPTIDES AND THEIR
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFOETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
; STREET: 1800 M Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-649-991-30

alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```
alignment_block:
US-09-332-522B-2 x US-08-649-991-30/rev ..
Align seg 1/1 to reverse of: US-08-649-991-30 from: 1 to: 54
973 PheSerThrSerArgAlaSer 980
|||||
26 TTTTCCACTTCCTCTAGAGCTTCT 3
seq_name: /cgn2_6/ptodata/2/1na/PCTUS_COMB.seq:PCT-US92-09202-2
seq_documentation_block:
; Sequence 2, Application PC/TUS9209202
; GENERAL INFORMATION:
; APPLICANT: Hogan, Michael E.
; TITLE OF INVENTION: Triplex Forming Oligonucleotide Reagents Targeted
; TITLE OF INVENTION: to the Neu Oncogene Promoter and Method of Use
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/09202
; FILING DATE: 19921028
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5393
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5151
; TELEFAX: 713/651-5246
; TELEX: 762829
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Triplex forming oligonucleotide
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
PCT-US92-09202-2

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x PCT-US92-09202-2/rev ..
Align seg 1/1 to reverse of: PCT-US92-09202-2 from: 1 to: 54
228 ProProHisHisHisHisHis 235
|||||
32 CCACCCACCACCCACCACCCAC 9
seq_name: /cgn2_6/ptodata/2/1na/5c_COMB.seq:US-08-332-766A-1
seq_documentation_block:
; Sequence 1, Application US/08332766A
; GENERAL INFORMATION:
; APPLICANT: PANNETIER, CHRISTOPHE
; APPLICANT: COCHET, MADELEINE
; APPLICANT: DARCHE, SYLVIE
; APPLICANT: KOURILSKY, PHILIPPE
; TITLE OF INVENTION: PROCESS FOR DETERMINING THE QUANTITY OF
; TITLE OF INVENTION: A DNA FRAGMENT OF INTEREST BY A METHOD OF
; TITLE OF INVENTION: METHOD OF ENZYMATIC AMPLIFICATION OF DNA
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
```

```
; Patent No. 5843647
; GENERAL INFORMATION:
; APPLICANT: JEFFREYS, Alec J.
; APPLICANT: ARMOUR, John
; TITLE OF INVENTION: SIMPLE TANDEM REPEATS
; NUMBER OF SEQUENCES: 125
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,766A
; FILING DATE: 01-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9326052.9
; FILING DATE: 21-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BIRD, Donald J.
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 217211/M94/0434/GB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-332-766A-1

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-08-332-766A-1/rev ..
Align seg 1/1 to reverse of: US-08-332-766A-1 from: 1 to: 377
228 ProProHisHisHisHisHis 235
|||||
202 CCACCCCATCACCACCATCAC 179
seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-263-413-23
seq_documentation_block:
; Sequence 23, Application US/08263413
; Patent No. 5747246
; GENERAL INFORMATION:
; APPLICANT: PANNETIER, CHRISTOPHE
; APPLICANT: COCHET, MADELEINE
; APPLICANT: DARCHE, SYLVIE
; APPLICANT: KOURILSKY, PHILIPPE
; TITLE OF INVENTION: PROCESS FOR DETERMINING THE QUANTITY OF
; TITLE OF INVENTION: A DNA FRAGMENT OF INTEREST BY A METHOD OF
; TITLE OF INVENTION: METHOD OF ENZYMATIC AMPLIFICATION OF DNA
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
```

```
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/263,413
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/882,980
; FILING DATE: 14-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavalleye, J.P.
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 660-058-55X
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-263-413-23

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-08-263-413-23 ..
Align seg 1/1 to: US-08-263-413-23 from: 1 to: 496

327 AspGluAspGluAspGluAsp 334
|||||
253 GATGAGGATGAGAGCAGGATG 276

seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-263-413-22

seq_documentation_block:
; Sequence 22, Application US/08263413
; Patent No. 5747246
; GENERAL INFORMATION:
; APPLICANT: PANNETIER, CHRISTOPHE
; APPLICANT: COCHET, MADELEINE
; APPLICANT: DARGHE, SYLVIE
; APPLICANT: KOURILSKY, PHILIPPE
; TITLE OF INVENTION: PROCESS FOR DETERMINING THE QUANTITY OF
; TITLE OF INVENTION: A DNA FRAGMENT OF INTEREST BY A METHOD OF
; TITLE OF INVENTION: METHOD OF ENZYMMATIC AMPLIFICATION OF DNA
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ORLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/263,413
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/882,980
; FILING DATE: 14-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavalleye, J.P.
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 660-058-55X
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-263-413-22

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-08-263-413-22 ..
Align seg 1/1 to: US-08-263-413-22 from: 1 to: 500

327 AspGluAspGluAspGluAsp 334
|||||
257 GATGAGGATGAGAGCAGGATG 280

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-07-807-043B-2

seq_documentation_block:
; Sequence 2, Application US/07807043B
; Patent No. 5342774
; GENERAL INFORMATION:
; APPLICANT: Boon, Thierry, Van den Eynde, Beno t
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; TITLE OF INVENTION: Rejection Antigens and Uses Thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/807,043B
; FILING DATE: 19911212
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
```



```
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA: 07/705,702
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5342774man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 253.3
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 675 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: singular
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-07-807-043B-2

alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-07-807-043B-2 ..
Align seg 1/1 to: US-07-807-043B-2 from: 1 to: 675

327 AspGluaspGluaspGluasp 334
|||||
247 GATGAGGATGAAGACGATGAGGAT 270

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-299-849B-2

seq_documentation_block:
; Sequence 2, Application US/08299849B
; Patent No. 5612201
; GENERAL INFORMATION:
; APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
; APPLICANT: Leth, Bernard; Szkora, Jean-Pierre; De Smet, Charles;
; APPLICANT: Chomez, Patrick
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
; TITLE OF INVENTION: Determining Expression Of A Tumor Antigen Precursor
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,849B
; FILING DATE: 1-SEPTEMBER-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
```

```
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA: 07/728,838
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5612201man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5355
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 675 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-299-849B-2

alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-08-299-849B-2 ..
Align seg 1/1 to: US-08-299-849B-2 from: 1 to: 675

327 AspGluaspGluaspGluasp 334
|||||
247 GATGAGGATGAAGACGATGAGGAT 270

seq_name: /cgn2_6/ptodata/2/1na/5D_COMB.seq:US-08-142-368A-2

seq_documentation_block:
; Sequence 2, Application US/08142368A
; Patent No. 5925729
; GENERAL INFORMATION:
; APPLICANT: Boon-Falleur, Thierry; Van der Bruggen, Thierry;
; APPLICANT: Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
; APPLICANT: Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; TITLE OF INVENTION: Rejection Antigens and Uses Thereof
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/142,368A
; FILING DATE: 02-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
```

```
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; APPLICATION NUMBER: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5925729man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5253.4-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 675 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-142-368A-2
```

```
alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
```

alignment_block:

US-09-332-522B-2 x US-08-142-368A-2 ..

Align seg 1/1 to: US-08-142-368A-2 from: 1 to: 675

```
327 AspGluAspGluAspGluasp 334
|||||
247 GATGAGGATGAAGACGATGAGGAT 270
```

seq_name: /cgn2_6/ptodata/2/ina/6-COMB.seq:US-08-967-727-2

seq_documentation_block:

; Sequence 2, Application US/08967727

; Patent No. 6025474

; GENERAL INFORMATION:

; APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;

; APPLICANT: van der Brugge, Pierre; Boon-Failleux, Thierry

; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For

; TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Felie & Lynch

; STREET: 805 Third Avenue

; CITY: New York City

; STATE: New York

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

; COMPUTER: IBM

; OPERATING SYSTEM: PC-DOS

; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/967,727

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/037,230

; FILING DATE: 26-MARCH-1993

; APPLICATION NUMBER: PCT/US92/04354

; FILING DATE: 22-MAY-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/807,043

; FILING DATE: 12-DECEMBER-1991

; PRIOR APPLICATION DATA:

```
; APPLICATION NUMBER: 07/764,365
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6025474man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5353
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 675 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-967-727-2
```

```
alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
```

alignment_block:

US-09-332-522B-2 x US-08-967-727-2 ..

Align seg 1/1 to: US-08-967-727-2 from: 1 to: 675

```
327 AspGluAspGluAspGluasp 334
|||||
247 GATGAGGATGAAGACGATGAGGAT 270
```

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5194596-8

seq_documentation_block:

; Patent No. 5194596

; APPLICANT: TISCHER, EDMUND G.; ABRAHAM, JUDITH A.; FIDDES, JOHN

; C.; MITCHELL, RICHARD L.

; TITLE OF INVENTION: PRODUCTION OF VASCULAR ENDOTHELIAL CELL

; GROWTH FACTOR

; NUMBER OF SEQUENCES: 32

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/450,883

; FILING DATE: 14-DEC-1989

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 387,545

; FILING DATE: 27-JUL-1989

; SEQ ID NO:8:

; LENGTH: 790

5194596-8

alignment_scores:

```
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
```

alignment_block:

US-09-332-522B-2 x 5194596-8 ..

Align seg 1/1 to: 5194596-8 from: 1 to: 790

```
227 ProProHisHisHisHisHis 234
|||||
416 CCACCACCACCACCACCACCAC 439
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-103-998-1

```
seq_documentation_block:
; Sequence 1, Application US/08103998
; Patent No. 5470725
; GENERAL INFORMATION:
; APPLICANT: Borriass, Rainer
; APPLICANT: Hofemeister, Jurgen
; APPLICANT: Thomsen, Karl Kristian
; APPLICANT: Olsen, Ole
; APPLICANT: Von Wetstein, Dietrich
; TITLE OF INVENTION: A Thermostable (1,3-1,4)-beta-Glucanase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOLEY & LARDNER
; STREET: 1800 Diagonal Road, P.O. Box 299
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/103,998
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/773,652
; FILING DATE: 16-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30307/123
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 469..1185
; US-08-103-998-1
```

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x US-08-103-998-1 ..

Align seg 1/1 to: US-08-103-998-1 from: 1 to: 1240

```
690 ThrGlyLeuPheMetSerLeuCys 697
|||||
496 ACCGATTGTTTATGAGTTGTGT 519
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-807-043B-4

```
seq_documentation_block:
; Sequence 4, Application US/07807043B
; Patent No. 5342774
; GENERAL INFORMATION:
; APPLICANT: Boon, Thierry, Van den Eynde, Beno t
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; TITLE OF INVENTION: Rejection Antigens and Uses Thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/807,043B
; FILING DATE: 19911212
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-May-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5342774man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 253.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-5200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1365 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: singular
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-07-807-043B-4
```

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x US-07-807-043B-4 ..

Align seg 1/1 to: US-07-807-043B-4 from: 1 to: 1365

```
327 AspGluAspGluAspGluasp 334
|||||
709 GATGAGGATGAAGACGATGAGGAT 732
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-299-849B-4

```
seq_documentation_block:
; Sequence 4, Application US/08299849B
; Patent No. 5612201
; GENERAL INFORMATION:
```

```
; APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
; APPLICANT: Leth, Bernard; Szikora, Jean-Pierre; De Smet, Charles;
; APPLICANT: Chomez, Patrick
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
```

```
; TITLE OF INVENTION: Determining Expression Of A Tumor Antigen Precursor
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,849B
; FILING DATE: 1-SEPTEMBER-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; APPLICATION NUMBER: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-May-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5612201man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5355
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1365 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-299-849B-4

alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-08-299-849B-4  ..

Align seg 1/1 to: US-08-299-849B-4 from: 1 to: 1365

327 AspGluAspGluAspGluAsp 334
|||||
709 GATGAGGATGAAGACGATGAGGAT 732

seq_name: /cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-142-368A-4

seq_documentation_block:
; Sequence 4, Application US/08142368A
; Patent No. 5925729
; GENERAL INFORMATION:
; APPLICANT: Boon-Falleur, Thierry; Van der Bruggen, Thierry;
```

```
; APPLICANT: Van den Eynde, Beno t; Van Pel, Allne; De Plaen, Etienne;
; APPLICANT: Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; TITLE OF INVENTION: Rejection Antigens and Uses Thereof
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/142,368A
; FILING DATE: 02-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; APPLICATION NUMBER: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-May-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5925729man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5253.4-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1365 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-142-368A-4

alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-08-142-368A-4  ..

Align seg 1/1 to: US-08-142-368A-4 from: 1 to: 1365

327 AspGluAspGluAspGluAsp 334
|||||
709 GATGAGGATGAAGACGATGAGGAT 732

seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-967-727-4

seq_documentation_block:
; Sequence 4, Application US/08967727
; Patent No. 6025474
; GENERAL INFORMATION:
; APPLICANT: Gaugier, B atrice; Van den Eynde, Beno t;
```

APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,727

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/037,230

FILING DATE: 26-MARCH-1993

APPLICATION NUMBER: PCT/US92/04354

FILING DATE: 22-MAY-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/807,043

FILING DATE: 12-DECEMBER-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/764,365

FILING DATE: 23-SEPTEMBER-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/728,838

FILING DATE: 9-JULY-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/705,702

FILING DATE: 23-MAY-1991

ATTORNEY/AGENT INFORMATION:

NAME: Hanson, No. 6025474man D.

REGISTRATION NUMBER: 30,946

REFERENCE/DOCKET NUMBER: LUD 5353

TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1365 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: genomic DNA

US-08-967-727-4

alignment_scores:

Quality: 8.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

Length: 8

Gaps: 0

alignment_block:

US-09-332-522B-2 x US-08-967-727-4

Align seg 1/1 to: US-08-967-727-4 from: 1 to: 1365

327 AspGluAspGluAspGluasp 334

|||||

709 GATGAGGATGAAGACGATGAGGAT 732

seq_name: /cgn2_5/ptodata/2/1na/5B_COMB.seq:US-08-485-284A-1

seq_documentation_block:

Sequence 1, Application US/08485284A

Patent No. 5750372

GENERAL INFORMATION:

APPLICANT: SAKAI, YASUYOSHI
APPLICANT: TANI, YOSHIKI
APPLICANT: SHIBANO, YUJI
APPLICANT: KONDO, HIROTO
APPLICANT: HATANAKA, HARUYO
TITLE OF INVENTION: NOVEL VECTOR HAVING PROMOTER THAT IS
INDUCIBLE BY METHANOL AND/OR GLYCEROL
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/485,284A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 43361/1992

FILING DATE: 28-FEB-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/025,416

FILING DATE: 01-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: WHITE JR., PAUL E.

REGISTRATION NUMBER: 32,011

REFERENCE/DOCKET NUMBER: 217755/FPS38209US

TELEPHONE: (202) 861-3000

TELEFAX: (202) 822-0944

TELEX: 6714627 CUSH

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1667 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: genomic DNA

US-08-485-284A-1

alignment_scores:

Quality: 8.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

Length: 8

Gaps: 0

alignment_block:

US-09-332-522B-2 x US-08-485-284A-1

Align seg 1/1 to: US-08-485-284A-1 from: 1 to: 1667

326 GlyAspGluAspGluAspGlu 333

|||||

1302 GGGGATGAAGATGAAGATGATGAG 1325

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-483-170-1

seq_documentation_block:

Sequence 1, Application US/08483170

Patent No. 5505943

GENERAL INFORMATION:

APPLICANT: Fortney, Donald Z.

APPLICANT: Durham, Donald D.

TITLE OF INVENTION: Compositions Containing Protease

TITLE OF INVENTION: Produced by Vibrio and Method of Use in

TITLE OF INVENTION: Debridement and Wound Healing

```
;
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W.R. Grace & Co. - Conn.
; STREET: 7379 Route 32
; CITY: Columbia
; STATE: Maryland
; COUNTRY: USA
; ZIP: 21044
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: DNASTAR
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,170
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/086,465
; FILING DATE:
; APPLICATION NUMBER: US 670612
; FILING DATE: 13-MAR-1991
; APPLICATION NUMBER: US 567884
; FILING DATE: 15-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Appleby, Vanessa L.
; REGISTRATION NUMBER: 33223
; REFERENCE/DOCKET NUMBER: 017914
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301 531 4515
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2000 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
;
; US-08-483-170-1
;
; alignment_scores:
; Quality: 8.00 Length: 8
; Ratio: 1.000 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 100.000
;
; alignment_block:
; US-09-332-522B-2 x US-08-483-170-1 rev ..
;
; Align seg 1/1 to reverse of: US-08-483-170-1 from: 1 to: 2000
;
; 639 TyrIleAlaArgArgSer 646
; |||||
; 1596 TACATCGCCAGCGGAGGAGT 1573
;
; seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-867-331-1
;
; seq_documentation_block:
; Sequence 1, Application US/08867331
; Patent No. 6017531
; GENERAL INFORMATION:
; APPLICANT: Fortney, Donald zane
; APPLICANT: Durham, Donald Richard
; APPLICANT: Yang, Kang
; TITLE OF INVENTION: HYDROPHILIC COMPOSITION CONTAINING
; TITLE OF INVENTION: PROTEASE PRODUCED BY VIBRIO
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. R. Grace & Co. -Conn.
; STREET: 7379 Route 32
; CITY: Columbia
; STATE: Maryland
; COUNTRY: United States
; ZIP: 21044
```

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,331
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 010440-068
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2000 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 61..1890
;
; US-08-867-331-1
;
; alignment_scores:
; Quality: 8.00 Length: 8
; Ratio: 1.000 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 100.000
;
; alignment_block:
; US-09-332-522B-2 x US-08-867-331-1/rev ..
;
; Align seg 1/1 to reverse of: US-08-867-331-1 from: 1 to: 2000
;
; 639 TyrIleAlaArgArgSer 646
; |||||
; 1596 TACATCGCCAGCGGAGGAGT 1573
;
; seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-942-001-1
;
; seq_documentation_block:
; Sequence 1, Application US/08942001A
; Patent No. 6020194
; GENERAL INFORMATION:
; APPLICANT: Mueller, Paul R.
; APPLICANT: Coleman, Thomas R.
; APPLICANT: Kumagai, Akiko
; APPLICANT: Dunphy, William G.
; TITLE OF INVENTION: Cdc2 Kinase, Myt 1, and Methods of Use
; FILE REFERENCE: 06618/074001
; CURRENT APPLICATION NUMBER: US/08/942,001A
; CURRENT FILING DATE: 1997-10-01
; EARLIER APPLICATION NUMBER: 60/028,073
; EARLIER FILING DATE: 1996-10-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2192
; TYPE: DNA
; ORGANISM: Xenopus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (169)..(1812)
;
; US-08-942-001-1
;
; alignment_scores:
; Quality: 8.00 Length: 8
```

```
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-332-522B-2 x US-08-942-001-1/rev ..
Align seg 1/1 to reverse of: US-08-942-001-1 from: 1 to: 2192

294 LeuLYSAsnLYsSerGIyAla 301
|||||
1860 TTGAATAAATAAAAGTCGTGCT 1837

seq_name: /cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-809-740A-1
seq_documentation_block:
; Sequence 1, Application US/08809740A
; Patent No. 5986077
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Process for producing anthracyclines
; and intermediates thereof
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Evenson, McKeown, Edwards and Lenahan
; STREET: 1200 G Street, Suite 700
; CITY: Washington
; STATE: DC
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,740A
; FILING DATE: 27-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FI95/00537
; FILING DATE: 30-SEP-1995
; APPLICATION NUMBER: FI 944556
; FILING DATE: 30-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: H. Thomas Anderton, Jr.
; REGISTRATION NUMBER: 40,895
; REFERENCE/DOCKET NUMBER: 1574/43419
; INFORMATION FOR SEQ ID NO: 1:
; LENGTH: 3252 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; STRAIN: Streptomyces nogalater ATCC 27451
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 359..1651
; OTHER INFORMATION: /note= "ORF1"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2937..3197
; OTHER INFORMATION: /note= "ORF2"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1648..1651
; OTHER INFORMATION: /note= "overlapping sequence in
; ORF1 and ORF2"
US-08-809-740A-1
```

```
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-332-522B-2 x US-08-809-740A-1/rev ..
Align seg 1/1 to reverse of: US-08-809-740A-1 from: 1 to: 3252

850 ValAspValValAspValSerArg 857
|||||
1294 GTTGACGTAGTCGACGTACGCCG 1271

seq_name: /cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-809-740A-4
seq_documentation_block:
; Sequence 4, Application US/08809740A
; Patent No. 5986077
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Process for producing anthracyclines
; and intermediates thereof
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Evenson, McKeown, Edwards and Lenahan
; STREET: 1200 G Street, Suite 700
; CITY: Washington
; STATE: DC
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,740A
; FILING DATE: 27-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FI95/00537
; FILING DATE: 30-SEP-1995
; APPLICATION NUMBER: FI 944556
; FILING DATE: 30-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: H. Thomas Anderton, Jr.
; REGISTRATION NUMBER: 40,895
; REFERENCE/DOCKET NUMBER: 1574/43419
; INFORMATION FOR SEQ ID NO: 4:
; LENGTH: 3252 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; STRAIN: Streptomyces nogalater ATCC 27451
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1648..2877
; OTHER INFORMATION: /note= "ORF2"
US-08-809-740A-4

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-332-522B-2 x US-08-809-740A-4/rev ..
```

Align seg 1/1 to reverse of: US-08-809-740A-4 from: 1 to: 3252

850 ValAspValValSerArg 857

|||||
1294 GTTCACGTAGTCGACGTCAGCGG 1271

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-07-807-043B-5

seq_documentation_block:

; Sequence 5, Application US/07807043B

; Patent No. 5342774

; GENERAL INFORMATION:

; APPLICANT: Boon, Thierry, Van den Eynde, Beno t

; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Felfe & Lynch

; STREET: 805 Third Avenue

; CITY: New York City

; STATE: New York

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

; COMPUTER: IBM

; OPERATING SYSTEM: PC-DOS

; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/807,043B

; FILING DATE: 19911212

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/764,364

; FILING DATE: 23-SEPTEMBER-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/728,838

; FILING DATE: 9-JULY-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/705,702

; FILING DATE: 23-May-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Hanson, No. 5342774man D.

; REGISTRATION NUMBER: 30,946

; REFERENCE/DOCKET NUMBER: LUD 253.3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 688-9200

; TELEFAX: (212) 838-3884

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4698 base pairs

; TYPE: NUCLEIC ACID

; STRANDEDNESS: singular

; TOPOLOGY: linear

; MOLECULE TYPE: genomic DNA

US-07-807-043B-5

alignment_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x US-07-807-043B-5 ..

Align seg 1/1 to: US-07-807-043B-5 from: 1 to: 4698

327 AspGluAspGluAspGluAsp 334

|||||

709 GATGAGGATGAAGACGATGAGGAT 732

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-299-849B-5

seq_documentation_block:

; Sequence 5, Application US/08299849B

; Patent No. 5612201

; GENERAL INFORMATION:

; APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;

; APPLICANT: Leth, Bernard; Szikora, Jean-Pierre; De Smet, Charles;

; APPLICANT: Chomez, Patrick

; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In

; NUMBER OF SEQUENCES: 48

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Felfe & Lynch

; STREET: 805 Third Avenue

; CITY: New York City

; STATE: New York

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

; COMPUTER: IBM

; OPERATING SYSTEM: PC-DOS

; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/299,849B

; FILING DATE: 1-SEPTEMBER-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/037,230

; FILING DATE: 26-MARCH-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/04354

; FILING DATE: 22-MAY-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/807,043

; FILING DATE: 12-DECEMBER-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/764,364

; FILING DATE: 23-SEPTEMBER-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/728,838

; APPLICATION NUMBER: 9-JULY-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/705,702

; FILING DATE: 23-May-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Hanson, No. 5612201man D.

; REGISTRATION NUMBER: 30,946

; REFERENCE/DOCKET NUMBER: LUD 5355

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 688-9200

; TELEFAX: (212) 838-3884

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4698 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: genomic DNA

US-08-299-849B-5

alignment_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x US-08-299-849B-5 ..

Align seg 1/1 to: US-08-299-849B-5 from: 1 to: 4698

327 AspGluAspGluAspGluAsp 334

|||||

709 GATGAGGATGAAGACGATGAGGAT 732

seq_name: /cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-142-368A-5

```
seq_documentation_block:
; Sequence 5, Application US/08142368A
; Patent No. 5925729
; GENERAL INFORMATION:
; APPLICANT: Boon-Falleur, Thierry; Van der Bruggen, Thierry;
; APPLICANT: Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
; TITLE OF INVENTION: Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felle & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/142,368A
; FILING DATE: 02-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; APPLICATION NUMBER:
; FILING DATE: 23-SEPTEMBER-1991
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; APPLICATION NUMBER: 07/728,838
; APPLICATION NUMBER: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-May-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5925729man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5253.4-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4698 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-142-368A-5
```

```
alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment_block:

US-09-332-522B-2 x US-08-142-368A-5 ..

Align seg 1/1 to: US-08-142-368A-5 from: 1 to: 4698

327 AspGluAspGluAspGluasp 334

|||||

709 GATGAGGATGAAGACGATGAGGAT 732

seq_name: /cgn2_5/ptodata/2/ina/6_COMB.seq:US-08-967-727-5

```
seq_documentation_block:
; Sequence 5, Application US/08967727
; Patent No. 6025474
; GENERAL INFORMATION:
; APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
; APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
; TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felle & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,727
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,365
; FILING DATE: 23-SEPTEMBER-1991
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6025474man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5353
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4698 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-967-727-5
```

```
alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment_block:

US-09-332-522B-2 x US-08-967-727-5 ..

Align seg 1/1 to: US-08-967-727-5 from: 1 to: 4698

327 AspGluAspGluAspGluasp 334

```
|||||
709 GATGAGATGAACGATGAGAT 732
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-131-365B-53
seq_documentation_block:
; Sequence 53, Application US/08131365B
; Patent No. 5527690
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/131,365B
; FILING DATE: 01-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-08-131-365B-53

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-08-131-365B-53 ..
Align seg 1/1 to: US-08-131-365B-53 from: 1 to: 5197

696 LeuCysAlaValAsnLeuAlaGlu 703
|||||
2200 TTGTGTGCGGTGAACCTGGCTGAA 2223

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-668-123-53
seq_documentation_block:
; Sequence 53, Application US/08668123
; Patent No. 5891631
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
```

```
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668,123
; FILING DATE: 14-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/131,365
; FILING DATE: 01-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-08-668-123-53

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x US-08-668-123-53 ..
Align seg 1/1 to: US-08-668-123-53 from: 1 to: 5197

696 LeuCysAlaValAsnLeuAlaGlu 703
|||||
2200 TTGTGTGCGGTGAACCTGGCTGAA 2223
```


THIS PAGE BLANK (USPTO)

CC represented.
SQ Sequence 1744 BP; 439 A; 352 C; 475 G; 478 T;

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x Q31638 ..

Align seg 1/1 to: Q31638 from: 1 to: 1744

258 ThrLysProThrHisLeuValGluPro 266

|||||

193 ACAAAACCAACCCACCTGTGGAGCCT 219

seq_name: N_Geneseq_36:Q31323

seq_documentation_block:

ID Q31323 standard; DNA; 1744 BP.

AC Q31323;

DT 20-APR-1993 (first entry)

DE E. brunetti ssRNA.

KW Small subunit; ribosomal RNA; amplification; PCR; Eimeria; ss.

PN EP-516385-A.

PD 02-DEC-1992.

PF 27-MAY-1992; 304781.

PR 29-MAY-1991; US-707362.

PR 12-MAY-1992; US-879469.

PA (MERI) MERCK & CO INC.

PI Chakraborty PR, Dashkevicz M, Elbrecht A, Feighner SD;

PI Liberator P, P-Juchelka H;

PI WPI; 92-400736/49.

DR Species-specific Eimeria tenella DNA probes - comprise divergent

PT DNA sequences and are complementary to E. tenella small sub-unit

PT ribosomal RNA gene

PS Disclosure; Fig 2; 79pp; English.

CC Comparative analysis of regions close to both the 3' and 5' ends

CC of small subunit ribosomal RNA sequences with near identity in the

CC eukaryotic kingdom identified two consensus sequences, ERIB 1 and

CC ERIB 10. ERIB 1 represents a consensus sequence less than 10

CC nucleotides from the 5' end of eukaryotic ssRNA genes. The ERIB 10

CC sequence is the inverse complement to a consensus sequence located

CC approx. 20 nucleotides from the 3' end of eukaryotic ssRNA genes.

CC Taken together, these two oligonucleotides span the vast majority of

CC the ssRNA gene sequence, and they were used as a primer pair in

CC PCR to selectively amplify the ssRNA genes contained within the

CC genomic DNA prep. from Eimeria brunetti. The ssRNA sequence

CC may then be used to design a species specific probes for diagnosis

CC of Eimeria brunetti infection. See also Q31283-332.

CC NOTE: As specifications EP-516381, EP-516383-6, EP-516391 and

CC EP-516395-6 are identical except in the claims section, sequences

CC for all these specifications can be found indexed under EP-516385.

CC However the claimed sequences of each specification will be indexed

CC under their own patent number, thus each separate patent will be

CC represented.

SQ Sequence 1744 BP; 439 A; 352 C; 475 G; 478 T;

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x Q31323 ..

Align seg 1/1 to: Q31323 from: 1 to: 1744

258 ThrLysProThrHisLeuValGluPro 266

|||||

193 ACAAAACCAACCCACCTGTGGAGCCT 219

seq_name: N_Geneseq_36:Q31470

seq_documentation_block:

ID Q31470 standard; DNA; 1744 BP.

AC Q31470;

DT 20-APR-1993 (first entry)

DE E. brunetti ssRNA.

KW Small subunit; ribosomal RNA; amplification; PCR; Eimeria; ss.

PN EP-516381-A.

PD 02-DEC-1992.

PF 27-MAY-1992; 304777.

PR 29-MAY-1991; US-707356.

PR 12-MAY-1992; US-879585.

PA (MERI) MERCK & CO INC.

PI Chakraborty PR, Dashkevicz M, Elbrecht A, Feighner SD;

PI Liberator P, P-Juchelka H;

PI WPI; 92-400732/49.

DR Species-specific Eimeria tenella DNA probes - comprise divergent

PT DNA sequences and are complementary to E. tenella small sub-unit

PT ribosomal RNA gene

PS Disclosure; Fig 2; 79pp; English.

CC Comparative analysis of regions close to both the 3' and 5' ends

CC of small subunit ribosomal RNA sequences with near identity in the

CC eukaryotic kingdom identified two consensus sequences, ERIB 1 and

CC ERIB 10. ERIB 1 represents a consensus sequence less than 10

CC nucleotides from the 5' end of eukaryotic ssRNA genes. The ERIB 10

CC sequence is the inverse complement to a consensus sequence located

CC approx. 20 nucleotides from the 3' end of eukaryotic ssRNA genes.

CC Taken together, these two oligonucleotides span the vast majority of

CC the ssRNA gene sequence, and they were used as a primer pair in

CC PCR to selectively amplify the ssRNA genes contained within the

CC genomic DNA prep. from Eimeria brunetti. The ssRNA sequence

CC may then be used to design a species specific probes for diagnosis

CC of Eimeria brunetti infection (method claimed). See also Q31465-72.

CC NOTE: As specifications EP-516381, EP-516383-6, EP-516391 and

CC EP-516395-6 are identical except in the claims section, sequences

CC for all these specifications can be found indexed under EP-516385.

CC However the claimed sequences of each specification will be indexed

CC under their own patent number, thus each separate patent will be

CC represented.

SQ Sequence 1744 BP; 439 A; 352 C; 475 G; 478 T;

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x Q31470 ..

Align seg 1/1 to: Q31470 from: 1 to: 1744

258 ThrLysProThrHisLeuValGluPro 266

|||||

193 ACAAAACCAACCCACCTGTGGAGCCT 219

seq_name: N_Geneseq_36:Q41906

seq_documentation_block:

ID Q41906 standard; DNA; 41 BP.

AC Q41906;

DT 30-SEP-1993 (first entry)

DE erb B2/neu promoter antiparallel TFO #2.

KW Parine; promoter; human; erb B2/neu; gene; HER-2; homologue; rat; neu;

KW triplex-forming oligonucleotide; TFO; retinoic acid; transgenic mice;

KW core promoter element; growth factor; c-AMP; cancer; mammary tumour;

KW tumour; NIH3T3 cells; pyrimidine; TPA; major groove; target; CAT box;

KW TATA box; transcription; transforming; AT box protein; RNA polymerase;

KW TFIID; control isomer; expression; ss.

OS Synthetic.

PN W09309788-A.
 PD 27-MAY-1993.
 PF 18-OCT-1992; U09202.
 PR 13-NOV-1991; US-792319.
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 PI HOGAN ME;
 DR WPI: 93-182231/22.
 PT Use of triplex-forming oligo-nucleotide - to inhibit
 PT proliferation of cells contg. an erb. B2/neu gene site, for
 PS treating cancers, psoriasis etc.
 PS Claim 6; Page 14; 26pp; English.
 CC The sequences given in Q41905-10 are triplex-forming oligonucleotides
 CC (TFO) which are specific to the promoter region of erb B2/neu. They
 CC bind to the major groove of the DNA duplex to form a triplex. The
 CC TFOs are complementary to the target sequence such they include a G
 CC when the complementary location in the DNA duplex is a GC pair and T
 CC when the complementary location in the duplex DNA is an AT base pair.
 CC The target site for these TFOs should have a stretch of DNA which is
 CC at least 65% purine or pyrimidine bases. The long purine run in the
 CC erb B2/neu promoter region includes the CAT box and the TATA box.
 CC Inhibition at the CAT box will inhibit transcription initiation by
 CC interfering directly with the CAT box protein-RNA polymerase
 CC interaction. Further inhibition of the protein binding at the CAT
 CC box site can also block the interaction of the CAT protein with TFIID
 CC at the TATA box. The sequences given in Q41911-13 are control isomers
 CC which comprise randomised sequences based on TFOs which do not bind to
 CC the erb B2/neu target sequence, and have no effect on erb B2/neu
 CC expression. The erb B2/neu (HER-2) gene is the human homologue of
 CC the rat neu gene. This human homologue is frequently amplified in
 CC tumours. When expressed at high levels in NIH3T3 cells, erb B2/neu
 CC is strongly transforming and results in a high incidence of mammary
 CC tumours in transgenic mice. The core promoter element of erb B2/neu
 CC resides within a 300 bp region of the 5' flanking domain. This region
 CC contains elements which confer sensitivity to enhance promoter
 CC function in the presence of cell growth factors such as TPA, c-AMP and
 CC retinoic acid. Therefore, overexpression of erb B2/neu may be one
 CC mechanism leading to cancer initiation or expression. Inhibition of
 CC the erb B2/neu promoter region by the TFOs may be used to inhibit
 CC expression of the gene and may therefore be used to treat or prevent
 CC cancers.
 SQ Sequence 41 BP; 0 A; 0 C; 25 G; 16 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x Q41906/rev ..
 Align seg 1/1 to reverse of: Q41906 from: 1 to: 41

228 ProProHisHisHisHisHisHis 235
 |||||||||||||||||||||
 32 CCACCCACCCACCCACCCACCCAC 9

seq_name: N_Geneseq_36:V55810
 seq_documentation_block:
 ID V55810 standard; DNA; 51 BP.
 AC V55810;
 DT 18-NOV-1998 (first entry)
 DE Chimeric construct GG24-N constructing primer 1.
 KW Fusion protein; stabilising polypeptide; proteolytic degradation;
 KW resistance; half-life; autoimmune disease; inflammation; nitro drug;
 KW IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;
 KW nitroreductase protein; enzyme therapy; prodrug therapy; protease;
 KW cancer; pathological condition; NTR gene; PCR primer; ss.
 OS Synthetic.
 PN W09822577-A1.
 PD 28-MAY-1998.
 PF 17-NOV-1997; IB1508.

PR 25-JUN-1997; US-048945.
 PR 13-NOV-1996; US-030986.
 PA (MASU/) MASUCCI M G.
 PI Masucci MG;
 DR WPI: 98-312463/27.
 PT New fusion proteins resistant to proteolytic degradation -
 PT comprising a core protein with a stabilising polypeptide comprising
 PT a peptide sequence containing glycine repeats
 PS Disclosure; Page 45; 120pp; English.
 CC Sequences shown in V55806 to V55811 represent primers used to construct
 CC chimeric constructs containing IkappaB gene and a stabilising sequence.
 CC The invention provides a method for increasing the resistance of a core
 CC protein to proteolytic degradation that comprises linking or inserting
 CC onto or into the core protein a stabilising polypeptide of formula
 CC (Glya)(Glyb)(Glyc)2n where Glya, Glyb, Glyc are 1-6 sequential Gly
 CC residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met, Phe, Pro or Thr
 CC and n can be anything between 1-66. X, Y and Z need not be identical from
 CC n repeat to n repeat. Alternatively a nucleic acid encoding a stabilising
 CC polypeptide can be linked onto or inserted into a nucleic acid encoding
 CC a core protein. The fusion proteins of the invention are more resistant
 CC to degradation by proteases and, thus, have a longer half-life than the
 CC unfused core protein. The products can be used for treating autoimmune
 CC diseases, cancer and inflammation. In particular, the core protein may
 CC be an IkappaB regulator protein for the treatment of inflammatory bowel
 CC disease, or a nitroreductase protein which can activate nitro drugs in
 CC enzyme/prodrug therapy to treat cancer or other pathological conditions.
 CC The fusion proteins can also be used in diagnostic methods such as in
 CC vivo imaging.
 SQ Sequence 51 BP; 2 A; 4 C; 34 G; 11 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x V55810/rev ..
 Align seg 1/1 to reverse of: V55810 from: 1 to: 51

228 ProProHisHisHisHisHisHis 235
 |||||||||||||||||||||
 46 CCCCCCACCACCCACCCACCCAC 23

seq_name: N_Geneseq_36:Q41905
 seq_documentation_block:
 ID Q41905 standard; DNA; 54 BP.
 AC Q41905;
 DT 30-SEP-1993 (first entry)
 DE erb B2/neu promoter antiparallel TPO #1.
 KW Purine; promoter; human; erb B2/neu; gene; HER-2; homologue; rat; neu;
 KW triplex-forming oligonucleotide; TPO; retinoic acid; transgenic mice;
 KW core promoter element; growth factor; c-AMP; cancer; mammary tumour;
 KW tumour; NIH3T3 cells; pyrimidine; TPA; major groove; target; CAT box;
 KW TATA box; transcription; transforming; AT box protein; RNA polymerase;
 KW TFIID; control isomer; expression; ss.
 OS Synthetic.
 PN W09309788-A.
 PD 27-MAY-1993.
 PF 28-OCT-1992; U09202.
 PR 13-NOV-1991; US-792319.
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 PI HOGAN ME;
 DR WPI: 93-182231/22.
 PT Use of triplex-forming oligo-nucleotide - to inhibit
 PT proliferation of cells contg. an erb. B2/neu gene site, for
 PT treating cancers, psoriasis etc.
 PS Claim 4; Page 14; 26pp; English.
 CC The sequences given in Q41905-10 are triplex-forming oligonucleotides
 CC (TFO) which are specific to the promoter region of erb B2/neu. They
 CC bind to the major groove of the DNA duplex to form a triplex. The

CC TFOS are complementary to the target sequence such they include a G
 CC when the complementary location in the DNA duplex is a GC pair and T
 CC when the complementary location in the duplex DNA is an A/T base pair.
 CC The target site for these TFOS should have a stretch of DNA which is
 CC at least 65% purine or pyrimidine bases. The long purine run in the
 CC erb B2/neu promoter region includes the CAT box and the TATA box.
 CC Inhibition at the CAT box will inhibit transcription initiation by
 CC interfering directly with the CAT box protein-RNA polymerase
 CC interaction. Further inhibition of the protein binding at the CAT
 CC box site can also block the interaction of the CAT protein with TFIIID
 CC at the TATA box. The sequences given in Q41911-13 are control isomers
 CC which comprise randomised sequences based on TFOS which do not bind to
 CC the erb B2/neu target sequence, and have no effect on erb B2/neu
 CC expression. The erb B2/neu (HER-2) gene is the human homologue of
 CC the rat neu gene. This human homologue is frequently amplified in
 CC tumours. When expressed at high levels in NIH3T3 cells, erb B2/neu
 CC is strongly transforming and results in a high incidence of mammary
 CC tumours in transgenic mice. The core promoter element of erb B2/neu
 CC resides within a 300 bp region of the 5' flanking domain. This region
 CC contains elements which confer sensitivity to enhance promoter
 CC function in the presence of cell growth factors such as TPA, c-AMP and
 CC retinoic acid. Therefore, overexpression of erb B2/neu may be one
 CC mechanism leading to cancer initiation or expression. Inhibition of
 CC the erb B2/neu promoter region by the TFOS may be used to inhibit
 CC expression of the gene and may therefore be used to treat or prevent
 CC cancers.
 SQ Sequence 54 BP; 0 A; 0 C; 28 G; 26 T;

alignment_scores: Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x Q41905/rev ..

Align seg 1/1 to reverse of: Q41905 from: 1 to: 54

228 ProProHSH1SH1SH1SH1S 235

|||||
 32 CCACCCACCCACCCACCCACCCAC 9

seq_name: N_Geneseq_36:T43662

seq_documentation_block:

ID T43662 standard; DNA; 54 BP.
 AC T43662;
 DT 19-AUG-1997 (first entry)
 DE HIV-1 matrix protein p17 gene fragment OSA.
 KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;
 KW detection; maternal transmission; hybridisation assay; immunoassay; ss.
 OS Human immunodeficiency virus type 1.
 FS key Location/Qualifiers
 FT mat_peptide 1..54
 FT /*tag= a
 FT /product= OSA_peptide
 FN EP-743364-A2.
 PD 20-NOV-1996.
 PF 17-MAY-1996; 401084.
 PR 18-MAY-1995; FR-005914.
 PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 FI Narwa R, Roques P;
 DR WPI; 96-507733/51.
 DR P-PSDB; W06619.

CC Human immunodeficiency virus p17 gene fragments, derived proteins
 CC and antibodies - useful for assessing the risk of maternal
 CC transmission of HIV-1 infection
 PS Claim 3; Page 28; 46pp; French.
 CC This sequence is a specifically claimed example of 21-90 nucleotide
 CC long nucleic acid fragments, derived from the gene encoding part of the
 CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at
 CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;

CC Y6 = ATA, TTA, CTG, CTA, CTA, CTG or ATG; Y7 = GAG or GAA; Y8 = GAA or
 CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or
 CC CAA; Y11 = AAA, AGT, AAG, AAT, AAC or ACC; provided that when Y5
 CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,
 CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.
 CC The new nucleic acid fragments can be used as reagents for determining
 CC and assessing the risk of maternal-fetal transmission of HIV-1, using
 CC standard hybridisation or immuno assays. The presence of such sequences
 CC in maternal blood is strongly correlated with transmission of infection.
 SQ Sequence 54 BP; 29 A; 7 C; 14 G; 4 T;

alignment_scores: Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x T43662/rev ..

Align seg 1/1 to reverse of: T43662 from: 1 to: 54

973 PheSerThrSerSerArgAlaSer 980

|||||
 26 TTTTCCACTTCTCTAGAGCTTCT 3

seq_name: N_Geneseq_36:V89537

seq_documentation_block:

ID V89537 standard; cDNA; 122 BP.
 AC V89537;
 DT 15-FEB-1999 (first entry)
 DE EST clone CQ309.
 KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
 KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
 KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
 KW gene therapy; ss.
 OS Homo sapiens.
 PN WO9845436-A2.
 PD 15-OCT-1998.
 PF 10-APR-1998; U06955.
 PR 10-APR-1997; US-838821.
 PA (GEMY) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, Lavalie ER, McCoy JM, Merberg D,
 PI Racie LA, Spaulding V, Treacy M;
 DR WPI; 99-070077/06.
 PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries.
 PS Claim 1; Page 249; 618pp; English.

CC The present sequence represents a human expressed sequence tag (EST).
 CC The polynucleotide, which is a secreted EST, and the encoded protein
 CC are predicted to have useful biological activities which would make
 CC them suitable for treating, preventing or ameliorating medical
 CC conditions in humans and animals, although no supporting data is
 CC given. Suggested activities include nutritional activity, immune
 CC stimulating or suppressing activity, haematopoiesis regulating
 CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
 CC activity, receptor/ligand activity, anti-inflammatory activity,
 CC cadherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The polynucleotide may also be useful for gene therapy.
 SQ Sequence 122 BP; 40 A; 22 C; 24 G; 36 T;

alignment_scores: Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x V89537 ..

Align seg 1/1 to: V89537 from: 1 to: 122

375 IleGlnLeuLysValLeuLeu 382
|||||
96 ATACAACAATTGAAGTGTACTC 119

seq_name: N_Geneseq_36:Q95175

seq_documentation_block:

ID Q95175 standard; DNA; 377 BP.
AC Q95175;
DT 08-FEB-1996 (first entry)
DE Simple tandem repeat (STR) corresponding to wg0e7.
KW Simple tandem repeat; STR; wg0e7; treatment; genetic; diagnosis;
KW characterisation; mapping; linkage studies; analysis; alleles; ss.
OS Synthetic.
PN WO9517522-A2.
PD 29-JUN-1995.
PF 21-DEC-1994; G02789.
PR 21-DEC-1993; GB-026052.
PA (UYLE-) UNIV LEICESTER.
PI Armour J, Jeffreys AJ;
DR WPI; 95-240682/31.
PT Identifying simple tandem repeat loci in DNA - by screening DNA
PT library to enrich for fragments contg. the repeats before cloning
PT and rescreening, also simple tandem repeats for treatment or
PT diagnosis
PS Claim 26; Page 17; 51pp; English.
CC Q95175 is a simple tandem repeat (STR) corresponding to wg0e7. The
CC STR can be used for treatment and diagnosis in human and veterinary
CC medicine, partic. for genetic characterisation, mapping, linkage
CC studies and analysis/diagnosis of acquired disease alleles.
SQ Sequence 377 BP; 70 A; 48 C; 143 G; 114 T;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x Q95175/rev ..

Align seg 1/1 to reverse of: Q95175 from: 1 to: 377

228 ProProHisHisHisHisHis 235
|||||
202 CCACCCCATCATCCACCACCATCAC 179

seq_name: N_Geneseq_36:V80689

seq_documentation_block:

ID V80689 standard; DNA; 433 BP.
AC V80689;
DT 26-MAR-1999 (first entry)
DE Melshan 3'-end of the porcine alpha-MSHR gene.
KW Porcine; wild boar; melshan; pietrain; large white; hamshire; duroc;
KW differentiation; breed origin; alpha-MSHR; coat colour; stock purity;
KW alpha melanocyte-stimulating hormone receptor; KIT; ss.
OS Sus scrofa.
PN WO9854360-A1.
PD 03-DEC-1998.
PF 27-MAY-1998; G01531.
PR 31-JAN-1998; GB-001990.
PR 30-MAY-1997; GB-011214.
PA (PIGL-) PIG IMPROVEMENT CO UK LTD.
PI Andersson L, Evans GJ, Giuffra E, Kijas J, Plastow GS,
PT Wales R;
DR WPI; 99-070222/06.
PT Differentiating products from different animal breeds - by the
PT analysis of alleles of breed-determinant genes, at the nucleic acid
PT or protein level
PS Claim 7; Fig 12; 101pp; English.

CC A method has been developed for: (a) differentiating animals and animal
CC products according to breed origin; (b) determining or testing the breed
CC origin of a product; or (c) validating an animal product. The method
CC comprises analysing a sample of the product for the allele(s) of at
CC least one breed-determinant (BD) gene. The present invention also
CC describes: (i) methods for determining the coat colour genotype of a pig
CC by determining; (ii) the allele(s) of the alpha melanocyte-stimulating
CC hormone receptor (alpha-MSHR) gene; (iii) the amino acid sequence of an
CC alpha-MSHR protein at positions associated with coat colour, or the size
CC of the protein; (iii) detecting which microsatellites (or other linked
CC marker alleles), linked to the alpha-MSHR gene, or particular alleles of
CC it, are present; and (iv) analysing nucleic acid to determine if the KIT
CC gene carries a polymorphism associated with the belt genotype. The
CC main method of the invention is applied to samples from fish, birds and
CC mammals, especially pigs. Particular applications are confirming stated
CC origin of meats; in quality control; for maintaining stock purity, and
CC in breeding programmes (to confirm particular crosses). The method
CC requires only very small samples and many samples can be screened
CC quickly and inexpensively. The process can be made quantitative. The
CC present sequence represents the 3'-end of the porcine alpha-MSHR gene
CC and the adjacent 3'-untranslated region from a Melshan breed pig.
SQ Sequence 433 BP; 63 A; 127 C; 177 G; 56 T;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x V80689 ..

Align seg 1/1 to: V80689 from: 1 to: 433

102 GlyGlyGlyArgGlyProSerLeu 109
|||||
397 GCGGAGGCGGGGCCCTCTT 420

seq_name: N_Geneseq_36:V80688

seq_documentation_block:

ID V80688 standard; DNA; 434 BP.
AC V80688;
DT 26-MAR-1999 (first entry)
DE Wild boar 3'-end of the porcine alpha-MSHR gene.
KW Porcine; wild boar; melshan; pietrain; large white; hamshire; duroc;
KW differentiation; breed origin; alpha-MSHR; coat colour; stock purity;
KW alpha melanocyte-stimulating hormone receptor; KIT; ss.
OS Sus scrofa.
PN WO9854360-A1.
PD 03-DEC-1998.
PF 27-MAY-1998; G01531.
PR 31-JAN-1998; GB-001990.
PR 30-MAY-1997; GB-011214.
PA (PIGL-) PIG IMPROVEMENT CO UK LTD.
PI Andersson L, Evans GJ, Giuffra E, Kijas J, Plastow GS,
PT Wales R;
DR WPI; 99-070222/06.
PT Differentiating products from different animal breeds - by the
PT analysis of alleles of breed-determinant genes, at the nucleic acid
PT or protein level
PS Claim 7; Fig 12; 101pp; English.
CC A method has been developed for: (a) differentiating animals and animal
CC products according to breed origin; (b) determining or testing the breed
CC origin of a product; or (c) validating an animal product. The method
CC comprises analysing a sample of the product for the allele(s) of at
CC least one breed-determinant (BD) gene. The present invention also
CC describes: (i) methods for determining the coat colour genotype of a pig
CC by determining; (ii) the allele(s) of the alpha melanocyte-stimulating
CC hormone receptor (alpha-MSHR) gene; (iii) the amino acid sequence of an
CC alpha-MSHR protein at positions associated with coat colour, or the size
CC of the protein; (iii) detecting which microsatellites (or other linked
CC marker alleles), linked to the alpha-MSHR gene, or particular alleles of

CC it, are present; and (iv) analysing nucleic acid to determine if the KIT
 CC gene carries a polymorphism associated with the Belt genotype. The
 CC main method of the invention is applied to samples from fish, birds and
 CC mammals, especially pigs. Particular applications are confirming stated
 CC origin of meats; in quality control; for maintaining stock purity, and
 CC in breeding programmes (to confirm particular crosses). The method
 CC requires only very small samples and many samples can be screened
 CC quickly and inexpensively. The process can be made quantitative. The
 CC present sequence represents the 3'-end of the porcine alpha-MSHR gene
 CC and the adjacent 3'-untranslated region from wild boar.
 SQ Sequence 434 BP; 65 A; 127 C; 176 G; 66 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x V80688 ..

Align seg 1/1 to: V80688 from: 1 to: 434

102 GlyGlyGlyArgGlyProSerLeu 109

|||||

398 GCGGAGGCGGGCGCCCTCTT 421

seq_name: N_Geneseq_36:V80690

seq_documentation_block:

ID V80690 standard; DNA; 434 BP.

AC V80690;

DE Hampshire 3'-end of the porcine alpha-MSHR gene.

KW Porcine; wild boar; meishan; pietrain; large white; hamshire; duroc;

KW Differentiation; breed origin; alpha-MSHR; coat colour; stock purity;

KW alpha melanocyte-stimulating hormone receptor; KIT; ss.

OS Sus scrofa.

PN WO984360-A1.

PD 03-DEC-1998.

PF 27-MAY-1998; G01531.

PR 31-JAN-1998; GB-001990.

PR 30-MAY-1997; GB-011214.

PA (FIG1-) PIG IMPROVEMENT CO UK LTD.

PI Andersson L, Evans GJ, Gluffa E, Kijas J, Plastow GS,

PI Wales R;

DR WPI; 99-070222/06.

PT Differentiating products from different animal breeds - by the

PT analysis of alleles of breed-determinant genes, at the nucleic acid

PT or protein level

PS Claim 7; Fig 12; 101pp; English.

CC A method has been developed for: (a) differentiating animals and animal
 CC products according to breed origin; (b) determining or testing the breed
 CC origin of a product; or (c) validating an animal product. The method
 CC comprises analysing a sample of the product for the allele(s) of at
 CC least one breed-determinant (BD) gene. The present invention also
 CC describes: (1) methods for determining the coat colour genotype of a pig
 CC by determining: (i) the allele(s) of the alpha melanocyte-stimulating
 CC hormone receptor (alpha-MSHR) gene; (ii) the amino acid sequence of an
 CC alpha-MSHR protein at positions associated with coat colour, or the size
 CC of the protein; (iii) detecting which microsatellites (or other linked
 CC marker alleles), linked to the alpha-MSHR gene, or particular alleles of
 CC it, are present; and (iv) analysing nucleic acid to determine if the KIT
 CC gene carries a polymorphism associated with the Belt genotype. The
 CC main method of the invention is applied to samples from fish, birds and
 CC mammals, especially pigs. Particular applications are confirming stated
 CC origin of meats; in quality control; for maintaining stock purity, and
 CC in breeding programmes (to confirm particular crosses). The method
 CC requires only very small samples and many samples can be screened
 CC quickly and inexpensively. The process can be made quantitative. The
 CC present sequence represents the 3'-end of the porcine alpha-MSHR gene
 CC and the adjacent 3'-untranslated region from a Hampshire breed pig.

CC Sequence 434 BP; 64 A; 127 C; 177 G; 66 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x V80690 ..

Align seg 1/1 to: V80690 from: 1 to: 434

102 GlyGlyGlyArgGlyProSerLeu 109

|||||

398 GCGGAGGCGGGCGCCCTCTT 421

seq_name: N_Geneseq_36:Q72471

seq_documentation_block:

ID Q72471 standard; cDNA; 675 BP.

AC Q72471;

DT 21-JUN-1995 (first entry)

DE Tumour rejection antigen MAGE-31 encoding cDNA.

KW Tumour rejection antigen; melanoma antigen-31; MAGE-31; MAGE-3;

KW cancer; cytolytic T cells; antigen D; human leucocyte antigen;

KW PIA gene; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT cds 1..675

FT /tag= a /

PN WO9423031-A.

PD 13-OCT-1994.

PF 17-MAR-1994; U02877.

PR 26-MAR-1993; US-037230.

PA (LUDWIG-) LUDWIG INST CANCER RES.

PI Boon-falleur T, Gaugler B, Van DEN EYNDE B, Van DER BRUGGEN P;

DR WPI; 94-333192/41.

DR P-PSDB; R74004.

PT New tumour rejection antigen precursor MAGE3 - useful in

PT treatment and diagnosis of cancer

PS Claim 4; Page 52; 105pp; English.

CC Q72471 is the cDNA sequence for the coding region of the gene

CC PIA, which encodes melanoma antigen-31 (MAGE-31, R74004). Also

CC encoded by the PIA gene is melanoma antigen-3 (MAGE-3, encoded

CC by Q72470) a tumour rejection antigen precursor, melanomas

CC characterised by the expression of MAGE-3 can be detected, or

CC monitored, by contacting a test sample with an agent that can

CC recognise MAGE-3. The melanoma can be treated by the administration

CC of cytolytic T cells specific for the complex of antigen D (the

CC mature rejection antigen derived from MAGE-3) and a human leucocyte

CC antigen (esp. HLA-A1).

SQ Sequence 675 BP; 200 A; 113 C; 214 G; 148 T;

alignment_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x Q72471 ..

Align seg 1/1 to: Q72471 from: 1 to: 675

327 AspGluAspGluAspGluAsp 334

|||||

247 GATGAGGATGAGGATGAGGAT 270

seq_name: N_Geneseq_36:T01174

seq_documentation_block:

ID T01174 standard; DNA; 675 BP.

AC T01174;

DT 26-FEB-1996 (first entry)
 DE P815A antigen precursor gene P1A exon 1.
 KW P815 antigen; P1A gene; cancer; diagnosis; ss.
 OS Mus musculus.
 PN WO9523874-A1.
 PD 08-SEP-1995.
 PF 23-FEB-1995; U02203.
 PR 01-MAR-1994; US-104727.
 PR 10-MAR-1994; US-209172.
 PR 01-SEP-1994; US-299849.
 PR 30-NOV-1994; US-346774.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Boon-Falleur T, Brasseur F, Chomez P, De Plaen E;
 PI De Smet C, Gaugler B, Lethe B, Marchand M, Fataad J;
 PI Szikora J, Van Den Eynde B, Van Derbruggen P, Weynants P;
 DR WPI; 95-320586/41.
 DR P-PSDB; R82989.
 PT Determ. of cancerous condition(s) - using a nucleic acid as a
 PT primer to determine expression of a MAGE tumour rejection antigen
 PT Precursor
 PS Disclosure; Page 66-67; 121pp; English.
 CC The coding sequence (T01174) of the P1A gene (T01176) encodes
 CC the P815A antigen precursor (R82989). The sequence was isolated
 CC in cosmid clone C1A.3.1 obtd. from DNA derived from mouse mastocytoma
 CC P815 cells.
 SQ Sequence 675 BP; 200 A; 113 C; 214 G; 148 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x T01174 ..
 Align seg 1/1 to: T01174 from: 1 to: 675

327 AspGluAspGluAspGluAsp 334
 |||||
 247 GATCAGCATGACGACGATGAGGAT 270

seq_name: N_Geneseq_36:T60966
 seq_documentation_block:
 ID T60966 standard; cDNA; 1023 BP.
 AC T60966;
 DT 23-JUN-1997 (first entry)
 DE NLERK2 cDNA.
 KW LERK; ligand for eph-related kinase; ERK; NLERK2;
 KW receptor protein tyrosine kinase; cell proliferation;
 KW cell differentiation; cell survival; nerve cell; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT signal_peptide 1..87
 FT mat_peptide 88..1020
 FT /tag= a
 FT /tag= b
 PN WO9704091-A1.
 PD 06-FEB-1997.
 PF 19-JUL-1996; AU0460.
 PR 20-JUL-1995; AU-004263.
 PR 27-NOV-1995; AU-006847.
 PR 22-DEC-1995; AU-007299.
 PR 05-FEB-1996; AU-007890.
 PA (AMRA-) AMRAD OPERATIONS PTY LTD.
 PI Nicola NA.
 DR WPI; 97-132632/12.
 DR P-PSDB; W10637.
 PT Nucleic acid mol. encoding ligand for eph-related kinase - useful
 PT for treatment of, pref. neuronal, cells to increase survival,
 PT proliferation and differentiation
 PS Claim 4; Page 37-39; 71pp; English.

CC A cDNA sequence (T60966) comprises the coding region for a novel
 CC ligand for eph-related kinase (LERK) designated NLERK2 (W10637).
 CC It was obtd. from a human foetal brain cDNA library using probes
 CC (see also T60967-69) based on an expressed sequence tag previously
 CC isolated on the basis of homology to conserved regions of known
 CC LERKs (see also W10633-36). NLERK2 3' (T60974) and 5' (T60975)
 CC regions were also detd. The isolated cDNA can be utilised in the
 CC prodn. of recombinant NLERK2 in transfected host cells for use e.g.
 CC in the treatment of conditions resulting from NLERK2 deficiency or
 CC defectiveness. Nucleic acids encoding NLERK2 may be introduced
 CC into a cell to enhance the ability of that cell to survive.
 SQ Sequence 1023 BP; 175 A; 306 C; 348 G; 194 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x T60966/rev ..
 Align seg 1/1 to reverse of: T60966 from: 1 to: 1023

101 LeuGlyGlyArgGlyProSer 108
 |||||
 671 CTGGGAGGGGCGAGGGGCTTCA 648

seq_name: N_Geneseq_36:V16097
 seq_documentation_block:
 ID V16097 standard; cDNA; 1120 BP.
 AC V16097;
 DT 06-JUL-1998 (first entry)
 DE Human transmembrane ligand Elk-L3 cDNA.
 KW Elk-L3; Elk-related receptor tyrosine kinase; transmembrane ligand;
 KW human; signal transduction; axonogenesis; nerve cell; neurone;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
 KW demyelination; multiple sclerosis; amyotrophic lateral sclerosis;
 KW nervous system infection; Wernicke's disease; trauma; ischaemia;
 KW stroke; nutritional polyneuropathy; progressive supranuclear palsy;
 KW Shy Drager's syndrome; multistem degeneration;
 KW olivo ponto cerebellar atrophy; peripheral nerve damage; ss.
 OS Chimeric - Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 60..1082
 FT /tag= a
 FT conflict 236
 FT /tag= b
 FT /note= "T in another cDNA clone"
 FT conflict 239
 FT /tag= c
 FT /note= "A in another cDNA clone"
 FT conflict 242
 FT /tag= d
 FT /note= "T in another cDNA clone"
 FT conflict 269
 FT /tag= e
 FT /note= "C in another cDNA clone"
 FT conflict 283
 FT /tag= f
 FT /note= "G in another cDNA clone"
 FT conflict 299
 FT /tag= g
 FT /note= "A in another cDNA clone"
 FT conflict 320
 FT /tag= h
 FT /note= "C in another cDNA clone"
 FT conflict 326
 FT /tag= i
 FT /note= "T in another cDNA clone"
 FT conflict 332

FT /tag- j
 FT /note- "I in another cDNA clone"
 FT conflict
 FT /tag- k
 FT /note- "T in another cDNA clone"
 FT conflict
 FT /tag- l
 FT /note- "A in another cDNA clone"
 FT conflict
 FT /tag- m
 FT /note- "C in another cDNA clone"
 FT conflict
 FT /tag- n
 FT /note- "G in another cDNA clone"
 FT conflict
 FT /tag- o
 FT /note- "C in another cDNA clone"
 FT conflict
 FT /tag- p
 FT /note- "A in another cDNA clone"
 FT conflict
 FT /tag- q
 FT /note- "C in another cDNA clone"
 FT conflict
 FT /tag- r
 FT /note- "C in another cDNA clone"
 FT conflict
 FT /tag- s
 FT /note- "A in another cDNA clone"
 FT conflict
 FT /tag- t
 FT /note- "A in another cDNA clone"
 FT conflict
 FT /tag- u
 FT /note- "C in another cDNA clone"
 FT conflict
 FT /tag- v
 FT /note- "A in another cDNA clone"
 FT conflict
 FT /tag- w
 FT /note- "A in another cDNA clone"
 FT conflict
 FT /tag- x
 FT /note- "A in another cDNA clone"
 FT conflict
 FT /tag- y
 FT /note- "T in another cDNA clone"
 FT conflict
 FT /tag- z
 FT /note- "G in another cDNA clone"
 FT conflict
 FT /tag- aa
 FT /note- "G in another cDNA clone"
 FT conflict
 FT /tag- ab
 FT /note- "A in another cDNA clone"
 FT conflict
 FT /tag- ac
 FT /note- "A in another cDNA clone"

WO9801548-A1.
 15-JAN-1998.
 04-JUL-1997; CA0473.
 05-JUL-1996; US-021272.
 (MOUN) MOUNT SINAI HOSPITAL CORP.
 Holland S, Mbamalu G, Pawson T;
 WPI: 98-101047/09.
 P-PSDB; W46615.
 DR Modulating transmembrane ligand for an Elk-related receptor tyrosine
 PT kinase - by formation of a complex between an oligomerised
 PT Elk-related receptor tyrosine kinase and a transmembrane ligand
 PS Disclosure; Fig 5A; 40pp; English.
 CC This cDNA sequence is a composite sequence of cDNAs encoding
 CC full-length human Elk-L3 and a segment of rat Elk-L3. Elk-L3 (see

CC W46615) is a transmembrane ligand of Elk-related receptor tyrosine
 CC kinase (ERRTK). A novel method of modulating the biological
 CC activity of, or affecting a pathway regulated by, a transmembrane
 CC ligand for an ERRTK in a cell expressing the transmembrane ligand
 CC comprises forming a complex between a purified and isolated
 CC oligomerised ERRTK, or an isoform or an extracellular domain of the
 CC ERRTK, and the transmembrane ligand expressed on the cell. The
 CC complex can also be used for evaluating a substance for its ability
 CC to modulate the biological activity of a transmembrane ligand for an
 CC ERRTK, and to identify substances that affect or modulate a pathway
 CC regulated by a ERRTK. A purified and isolated oligomerised ERRTK can
 CC be used in the preparation of a medicament for modulating neuronal
 CC development or regeneration in a subject, or in a medicament for
 CC modulating axonogenesis in a subject (all claimed). The substances
 CC identified by the methods can be used to modulate axonogenesis,
 CC nerve cell interactions and regeneration, to treat diseases and
 CC conditions involving trauma and injury to the nervous system, e.g.
 CC Alzheimer's disease, Parkinson's disease, Huntington's disease,
 CC demyelinating diseases, such as multiple sclerosis, amyotrophic
 CC lateral sclerosis, bacterial and viral infections of the nervous
 CC system, deficiency diseases, such as Wernicke's disease and
 CC nutritional polyneuropathy, progressive supranuclear palsy,
 CC Shy Drager's syndrome, multistem degeneration and olivoponto
 CC cerebellar atrophy, peripheral nerve damage, trauma, and ischaemia
 CC resulting from stroke. 181 A; 334 C; 390 G; 215 T;
 SQ Sequence 1120 BP;

alignment_scores:

Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x V16097/rev ..

Align seg 1/1 to reverse of: V16097 from: 1 to: 1120

101 LeuGlyGlyArgGlyProser 108

730 CTGGGAGGGCGAGGGGCGCTTCA 707

seq_name: N_Geneseq_36.Q05832

seq_documentation_block:

ID Q05832 standard; DNA; 1240 BP.
 AC Q05832;
 DT 09-JAN-1991 (first entry)
 DE Beta-glucanase-H1 gene.
 KW Hybrid pre-beta-glucanase; glucans; beer; animal feed; poultry; ss.
 OS Bacillus amyloliquefaciens, Bacillus macerans.
 FH Key Location/Qualifiers
 cds 469..1188
 /tag- a
 /label-Hybrid pre-glucanase
 FT misc_rna 469..864
 /tag- b
 /label-N-terminal of B.amyloliquefaciens beta-
 FT glucanase
 FT 865..1185
 /tag- c
 /label-C-terminal half of B.macerans beta-
 FT glucanase
 FT
 FT WO9009436-A.
 FT 23-AUG-1990.
 FT 16-FEB-1990; DK0044.
 FT 16-FEB-1989; DD-325800.
 FT 04-AUG-1989; DK-003848.
 FT (CARL-) CARLSBERG A/S.
 FT (DEAK) AKAD WISSENSCHAFT DDR.
 FT Borriass R, Hofemeister J, Thomsen KK, Olsen O, Vonvettstein D;
 DR WPI: 90-275129/36.
 DR P-PSDB; R06621.

PT New thermostable (1,3-1,4)-beta-glucanase - prepd. using hybrid
 PT gene obtd. using Bacillus amyloliquefaciens and B.macerans genes
 PS Disclosure; page 26; 84pp; English.
 CC This beta-glucanase-HI gene encodes a hybrid pre-beta-glucanase
 CC comprising a signal peptide and the amino-terminal of the B.
 CC amyloliquefaciens beta-glucanase and the carboxyl-terminal half of
 CC the B.macerans beta-glucanase. The hybrid (1,3-1,4)-beta-gluc-
 CC anase-HI encoded is thermostable and hydrolyses beta-glycosidic
 CC linkages in (1,3-1,4)-beta-glucans. Reducing sugars are obtd. at
 CC high temps. and thus this enzyme can be used in the mfr. of food
 CC prods., esp. beer and animal feed (e.g. for feeding poultry).
 CC See also Q05833.
 SQ Sequence 1240 BP; 370 A; 223 C; 287 G; 360 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x Q05832 ..
 Align seg 1/1 to: Q05832 from: 1 to: 1240

690 ThrGlyLeuPheMetSerLeuCyS 597
 |||||
 496 ACCGGATTGTTTATGAGTTGTGT 519

seq_name: N_Geneseq_36:Q32350

seq_documentation_block:
 ID Q32350 standard; cDNA; 1365 BP.
 AC Q32350;
 DT 22-APR-1993 (first entry)
 DE PIA cDNA.
 KW Tumour rejection antigen precursor; TRAP; PIA; Northern blot; PCR;
 KW polymerase chain reaction; P1.HTR; P815A; ORF; promoter; CAAT box;
 KW enhancer; MHC; class I; mouse; nucleolar protein; NO38; B23;
 KW nuclear targeting site; acidic domain; phosphorylation site;
 KW serine; glutamate; P91A; 35B; P198; antigen; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 463..1137
 FT /*tag= a

WO9220356-A.
 26-NOV-1992.
 22-MAY-1992; U04354.
 23-MAY-1991; US-705702.
 09-JUL-1991; US-728838.
 23-SEP-1991; US-764364.
 12-DEC-1991; US-807043.
 (LUDW-) LUDWIG INST CANCER RES.
 PI Boon T, Chomez P, De Plaen E, Lurquin C, Traversari C;
 PI Van Den Eynde B, Van Der Bruggen P, Van Pel A;
 DR WPI; 92-41540/50.
 DR Nucleic acid mol. encoding a human tumour rejection antigen
 PT precursor - useful as an immunostimulant in a vaccine for
 PT treating and preventing cancers, also useful in diagnosis
 PS Disclosure; page 65; 142pp; English.
 CC This sequence represents the cDNA of the tumour rejection antigen
 CC precursor (TRAP), PIA. This sequence was isolated by Northern blot
 CC analysis and polymerase chain reaction from the cell line P1.HTR. The
 CC coding sequence depicts the exons from gene P815A. This sequence was
 CC then used to isolate the P815A full length coding sequence (see also
 CC Q32370). The full length gene contains three exons with the ORF for
 CC the 244 amino acid PIA protein starting in exon 1 and terminating in
 CC exon 2. The promoter region contains a CAAT box and an enhancer
 CC sequence, this being similar to those found in the promoters of
 CC most MHC class I genes. One region of 95 bases within exon 1 was
 CC found to bear homology to acidic regions in mouse nucleolar protein
 CC NO38/P23, 56/95 of the bases were identical. The coded product of
 CC this gene is thought to have a molecular mass of 25 kD. This gene is

CC thought to have a potential nuclear targeting site at residues 5-9,
 CC as well as a large acidic domain at position 83-118. A putative
 CC phosphorylation site can be found at position 125 (serine) and a
 CC second acidic domain is found close to the C-terminus as an
 CC uninterrupted stretch of 14 glutamate residues. The PIA gene has
 CC been found to be dissimilar to the sequences for P91A, 35B and P198,
 CC showing that PIA is indicative of a different class of genes and
 CC antigens. 1365 BP; 373 A; 277 C; 362 G; 353 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x Q32350 ..
 Align seg 1/1 to: Q32350 from: 1 to: 1365

327 AspGluAspGluAspGluasp 334
 |||||
 709 GATGAGGTGAAGACGATGAGGAT 732

seq_name: N_Geneseq_36:Q72474

seq_documentation_block:
 ID Q72474 standard; cDNA; 1365 BP.
 AC Q72474;
 DT 21-JUN-1995 (first entry)
 DE Entire PIA cDNA.
 KW Tumour rejection antigen; melanoma antigen-31; MAGE-31; MAGE-3
 KW cancer; cytolytic T cells; antigen D; human leucocyte antigen;
 KW PIA gene; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 463..1137
 FT /*tag= a

WO9423031-A.
 13-OCT-1994.
 17-MAR-1994; U02877.
 26-MAR-1993; US-037230.
 (LUDW-) LUDWIG INST CANCER RES.
 PI Boon-falleur T, Gaugler B, Van DEN EYNDE B, Van DER BRUGGEN P;
 DR WPI; 94-333192/41.
 DR New tumour rejection antigen precursor MAGE3 - useful in
 PT treatment and diagnosis of cancer
 PS Disclosure; Page 54; 105pp; English.
 CC Q72473 is the entire PIA cDNA containing the coding region
 CC (Q72471), which encodes melanoma antigen-31 (MAGE-31, R74004).
 CC Also encoded by the PIA gene is melanoma antigen-3 (MAGE-3, encoded
 CC by Q72470) a tumour rejection antigen precursor, melanomas.
 CC characterised by the expression of MAGE-3 can be detected, or
 CC monitored, by contacting a test sample with an agent that can
 CC recognise MAGE-3. The melanoma can be treated by the administration
 CC of cytolytic T cells specific for the complex of antigen D (the
 CC mature rejection antigen derived from MAGE-3) and a human leucocyte
 CC antigen (esp. HLA-A1).
 SQ Sequence 1365 BP; 373 A; 277 C; 362 G; 353 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x Q72474 ..
 Align seg 1/1 to: Q72474 from: 1 to: 1365

327 AspGluAspGluAspGluasp 334


```

|||||
709 GATGAGGATGAACGATGAGGAT 732
seq_name: N_Geneseq_36:T05084

seq_documentation_block:
ID T05084 standard; DNA; 1365 BP.
AC T05084;
DT 29-FEB-1996 (first entry)
DE P815A antigen precursor gene p1A.
KW P815 antigen; p1A gene; cancer; diagnosis; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT cds 463..1137
FT /*tag= a
FT /product= P815 antigen precursor
PN WO9523874-A1.
PD 08-SEP-1995.
PF 23-FEB-1995; U02203.
PR 01-MAR-1994; US-204727.
PR 10-MAR-1994; US-209172.
PR 01-SEP-1994; US-299849.
PR 30-NOV-1994; US-346774.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Boon-Falleur T, Brasseur F, Chomez P, De Plaen B;
PI De Smet C, Gaugler B, Lethe B, Marchand M, Fataud J;
PI Szikora J, Van Den Eynde B, Van Derbruggen P, Weynants P;
DR WPI; 95-320586/41.
PT Determ. of cancerous condition(s) - using a nucleic acid as a
PT precursor
PS Example 9; Page 67; 121pp; English.
CC The p1A gene and 3' and 5' flanking sequences (T05084) was isolated
CC from cosmid clone CIA.3.1 obtd. from DNA derived from mouse
CC mastocytoma P815 cells. The gene codes for P815A antigen precursor
CC (R82989).
SQ Sequence 1365 BP; 373 A; 277 C; 362 G; 353 T;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x T05084 ..
Align seg 1/1 to: T05084 from: 1 to: 1365

327 AspGluaspGluaspGluasp 334
|||||
709 GATGAGGATGAACGATGAGGAT 732
seq_name: N_Geneseq_36:T14691

seq_documentation_block:
ID Q48589 standard; DNA; 1667 BP.
AC Q48589;
DT 16-FEB-1994 (first entry)
DE Alcohol oxidase gene promoter.
KW Expression cassette; expression vector; inducible; methanol;
KW glycerol; yeast; ss.
OS Saccharomyces cerevisiae.
FH Key Location/Qualifiers
FT tata_signal 1329..1335
FT /*tag= a
PN EP-558024-A.
PD 01-SEP-1993.
PF 26-FEB-1993; 103040.
PR 28-FEB-1992; JP-043361.
PA (SUNR) SUNTORY LTD.
PI Hatanaka H, Kondo H, Sakai Y, Shibano Y, Tani Y;
DR WPI; 93-274483/35.
PT New expression cassettes inducible by methanol and/or glycerol -
PT having promoter and terminator derived from alcohol oxidase gene
PS Claim 1; Page 14-15; 37pp; English.
CC An expression cassette which contains the alcohol oxidase promoter
CC is induced when in the presence of methanol/glycerol. Under such
CC conditions any heterologous gene located downstream of the promoter
CC will be expressed. The expression cassette can be used to construct
CC an expression vector which in turn can be used to transform cells.
CC When transformed cells are cultured in the presence of methanol/
CC glycerol expression of the heterologous gene is induced such that
CC large quantities of the desired product is produced. The
CC transformed cells are used particularly for the production of
CC enzymes such as adenylate kinases, cytochrome C and peroxidases.
SQ Sequence 1667 BP; 546 A; 276 C; 204 G; 641 T;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x Q48589 ..
Align seg 1/1 to: Q48589 from: 1 to: 1667

```

```

DR P-PSDB; R94615.
PT Immunogenic compositions containing tumour rejection antigen
PT precursors - to stimulate immune response for treatment of tumours
PS Claim 7; Page 21-22; 90pp; English.
CC MAGE tumour rejection antigen precursor (TRAP) DNA (T14691) and
CC genomic DNA (T14692) are a source of the isolated and purified TRAP
CC (see also R94515) and the tumour rejection antigen, either of which
CC can be used as an agent for the treatment of the cancer for which
CC the antigen is a marker, as well as in various diagnostic and
CC surveillance approaches to oncology.
SQ Sequence 1365 BP; 373 A; 277 C; 362 G; 353 T;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x T14691 ..
Align seg 1/1 to: T14691 from: 1 to: 1365

327 AspGluaspGluaspGluasp 334
|||||
709 GATGAGGATGAACGATGAGGAT 732
seq_name: N_Geneseq_36:Q48589

seq_documentation_block:
ID Q48589 standard; DNA; 1667 BP.
AC Q48589;
DT 16-FEB-1994 (first entry)
DE Alcohol oxidase gene promoter.
KW Expression cassette; expression vector; inducible; methanol;
KW glycerol; yeast; ss.
OS Saccharomyces cerevisiae.
FH Key Location/Qualifiers
FT tata_signal 1329..1335
FT /*tag= a
PN EP-558024-A.
PD 01-SEP-1993.
PF 26-FEB-1993; 103040.
PR 28-FEB-1992; JP-043361.
PA (SUNR) SUNTORY LTD.
PI Hatanaka H, Kondo H, Sakai Y, Shibano Y, Tani Y;
DR WPI; 93-274483/35.
PT New expression cassettes inducible by methanol and/or glycerol -
PT having promoter and terminator derived from alcohol oxidase gene
PS Claim 1; Page 14-15; 37pp; English.
CC An expression cassette which contains the alcohol oxidase promoter
CC is induced when in the presence of methanol/glycerol. Under such
CC conditions any heterologous gene located downstream of the promoter
CC will be expressed. The expression cassette can be used to construct
CC an expression vector which in turn can be used to transform cells.
CC When transformed cells are cultured in the presence of methanol/
CC glycerol expression of the heterologous gene is induced such that
CC large quantities of the desired product is produced. The
CC transformed cells are used particularly for the production of
CC enzymes such as adenylate kinases, cytochrome C and peroxidases.
SQ Sequence 1667 BP; 546 A; 276 C; 204 G; 641 T;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x Q48589 ..
Align seg 1/1 to: Q48589 from: 1 to: 1667

```

326 GlyAspGluAspGluAspGlu 333
 1302 GGGGATGAAGATGAAGATGAG 1325

seq_name: N_Geneseq_36:V17568

seq_documentation_block:

ID V17568 standard; cDNA; 1689 BP.
 AC V17568;
 DT 10-JUN-1998 (first entry)
 DE Coding sequence for Gv4 subunit of glycinin.
 KW Beta-conglycinin; soybean seed protein; transgenic plant; glycinin;
 KW seed storage protein profile; ss.
 OS Glycine max.
 PN WO9747731-A2.
 PD 18-DEC-1997.
 PF 14-JUN-1997; U09743.
 PR (DUPO) DU PONT DE NEMOURS & CO E I.
 PA Fader GM, Kinney AJ;
 PI Wri; 98-052298/05.
 DR
 PT Suppression of specific classes of soybean seed protein genes -
 PT useful to change seed storage protein profiles of transgenic plants
 PS Disclosure; Page 37-38; 58pp; English.
 CC This sequence represents the coding sequence for the Gv4 subunit of the
 CC soybean seed protein, glycinin. The method of the invention is for
 CC reducing the quantity of a soybean seed storage protein (A), such as
 CC beta-conglycinin, in soybeans. It comprises: (a) constructing a chimeric
 CC gene comprising: (i) a nucleic acid fragment encoding a promoter that is
 CC functional in the cells of soybean seeds; (ii) a nucleic acid fragment
 CC encoding all or a portion of (A) placed in sense or antisense orientation
 CC relative to the promoter of (i); and (iii) a transcriptional termination
 CC region; (b) creating a transgenic soybean cell by introducing into a
 CC soybean cell the chimeric gene of (a); and (c) growing the transgenic
 CC soybean cells of (b) under conditions that result in expression of the
 CC chimeric gene of (a); where the quantity of one or more members of a
 CC class of (A) subunits is reduced when compared to soybeans not containing
 CC the chimeric gene of (a). The method is used to construct transgenic
 CC soybean lines where the expression of genes encoding (A) are modulated to
 CC effect a change in seed storage protein profile of transgenic plants.
 CC Modification of the seed storage protein profile can result in the
 CC production of novel soy protein products with unique and valuable
 CC functional characteristics.
 SQ Sequence 1689 BP; 515 A; 433 C; 404 G; 337 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x V17568 ..

Align seg 1/1 to: V17568 from: 1 to: 1689

327 AspGluAspGluAspGluAsp 334
 850 GATGAAGACGAAGATGATGAAGAT 873

seq_name: N_Geneseq_36:T89519

seq_documentation_block:

ID T89519 standard; cDNA; 1708 BP.
 AC T89519;
 DT 14-APR-1998 (first entry)
 DE Human cytokine Lerk-8 cDNA.
 KW Lerk-8; cytokine; human; hek; elk; receptor tyrosine kinase;
 KW ligand; neurodegenerative disease; wound healing;
 KW neovascularisation; diagnosis; therapy; ss.
 OS Homo sapiens.
 FH Key
 FT CDS
 Location/Qualifiers
 398..1420

FT sig_peptide /*tag= a
 FT 398..478 /*tag= b
 FT 479..1417 /*tag= c
 FT mat_peptide
 FT variation 1370
 FT /*tag= d
 FT /*note= "nucleotide 1370 is C in one isolated
 FT cDNA clone (alters Val-298 codon to Leu)"
 PN WO9736919-A2.
 PD 09-OCT-1997.
 PF 19-MAR-1997; U04533.
 PR 21-MAR-1996; US-621146.
 PA (IMM) IMMUNEX CORP.
 PI Cerretti DP;
 PD WPI; 97-503043/46.
 DR P-PSDB; W31544.
 PT New isolated cytokine, Lerk-8 - binds to the hek and elk receptor
 PT tyrosine kinases, used to develop products for diagnosis and therapy
 PS Claim 3; Page 30-32; 37pp; English.
 CC This cDNA clone codes for a novel human cytokine designated Lerk-8
 CC (see W31544). Lerk-8 binds to the cell surface receptors hek and
 CC elk, which are members of the eph/elk family of receptor tyrosine
 CC kinases. The clone was isolated from a human foetal brain library
 CC using as probe a PCR fragment obtained using primers based on an
 CC EST (Genbank Acc. No. H10006) identified as having homology to
 CC Lerk-2 and Lerk-8 sequences. Lerk-8 polynucleotides can be used to
 CC prepare recombinant Lerk-8 polypeptides, especially soluble
 CC polypeptides comprising the extracellular domain of Lerk-8, in host
 CC cells. These polypeptides can be used to purify hek or elk proteins,
 CC and to purify or identify cells that express hek or elk on the
 CC surface. Such cells can be used in various in vitro studies or in
 CC vivo procedures, e.g. neural cells expressing elk can be administered
 CC to a mammal afflicted with a neurodegenerative disorder. The Lerk-8
 CC polypeptides can also be used to deliver diagnostic or therapeutic
 CC agents to these cells (e.g. leukaemia cells). The Lerk-8 DNA and
 CC polypeptides can also be used to treat disorders mediated by
 CC defective or insufficient amounts of Lerk-8; to treat disorders
 CC such as injury to neural tissue or neurologic disease; to promote
 CC angiogenesis; and for wound healing or stimulating
 CC neovascularisation of grafted tissues.
 SQ Sequence 1708 BP; 244 A; 563 C; 572 G; 329 T;

alignment_scores:

Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x T89519/rev ..

Align seg 1/1 to reverse of: T89519 from: 1 to: 1708

101 LeuClGlyGlyArgGlyProSer 108
 1068 CTGGAGGGGGGAGGGGGCTTCA 1045

seq_name: N_Geneseq_36:X13178

seq_documentation_block:

ID X13178 standard; DNA; 1814 BP.
 AC X13178;
 DT 19-MAR-1999 (first entry)
 DE Enterococcus faecalis genome contig SEQ ID NO:241.
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;
 KW vaccine; attenuation; computer readable medium; ds.
 OS Enterococcus faecalis.
 PN WO980555-A2.
 PD 12-NOV-1998.
 PF 04-MAY-1998; U08985.
 PR 14-NOV-1997; US-066009.
 PO 06-MAY-1997; US-044031.

PR 16-MAY-1997; US-046655.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Barash SC, Dillon PJ, Kunsch CA;
 DR WPI; 99-045171/04.
 FT New isolated Enterococcus faecalis polynucleotides and polypeptides
 FT - used to develop products for the detection of Enterococcus and for
 FT use in vaccines for prevention or attenuation of Enterococcus
 FT infection.
 PS Claim 1: Page 1201-1202; 2084pp; English.
 CC A computer readable medium has been developed which has recorded on it
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
 CC X12938 to X13919 represent these nucleotide sequences which are primary
 CC nucleotide sequences, also known as contigs. The computer-based system
 CC can identify fragments of the Enterococcus faecalis genome with
 CC commercial importance. The products can be used to detect the presence
 CC of Enterococcus faecalis in samples. They can also be used for
 CC diagnosing Enterococcal infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or
 CC another related organism, in vivo or in vitro. In particular the
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an Enterococcal
 CC infection.
 SQ Sequence 1814 BP; 603 A; 266 C; 277 G; 661 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x X13178 ..
 Align seg 1/1 to: X13178 from: 1 to: 1814

788 LysGlnLysGlnLeuArgPheGly 795
 ||||||||||||||||||||
 28 AACAAACAACTAGATTCGCT 51

seq_name: N_Geneseq_36:T69808

seq_documentation_block:
 ID T69808 standard; DNA; 1860 BP.
 AC T69808;
 DE 09-AUG-1997 (first entry)
 DE EPH family ligand Efl-6 gene.
 KW Efl-6; Eph; Elk; receptor tyrosine kinase; signal transduction;
 KW ligand; neurological disease; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 202..1224
 FT /*tag= a
 FT signal_peptide 202..273
 FT /*tag= b
 FT mat_peptide 274..1221
 FT /*tag= c
 FT misc_difference 65
 FT /*tag= d
 FT /*note= "base 65 is given as r in the specification"
 FT misc_difference 698
 FT /*tag= e
 FT /*note= "base 698 is given as A/G in the
 FT specification"
 FT misc_difference 1717
 FT /*tag= f
 FT /*note= "base 1717 is given as n in the
 FT specification"
 FT misc_difference 1741
 FT /*tag= g
 FT /*note= "base 1742 is given as n in the
 FT specification"
 FT misc_difference 1782

FT /*tag= h
 FT /*note= "base 1782 is given as n in the
 FT specification"
 FT misc_difference 1795
 FT /*tag= i
 FT /*note= "base 1795 is given as n in the
 FT specification"
 PN WO9715667-A1.
 PD 01-MAY-1997.
 PR 25-OCT-1996; U17201.
 PR 25-OCT-1996; US-007015.
 PA (REG-) REGENERON PHARM INC.
 PI Davis S, Gale NW, Yancopoulos GD;
 DR WPI: 97-259021/23.
 DR P-PSDB; W17081.
 CC New nucleic acid encoding Efl-6 ligand protein - used for promoting
 CC growth and proliferation of neuronal cells and in drug screening
 CC Claim 1; Fig 1; 36pp; English.
 CC An isolated DNA molecule (T69808) codes for a novel ligand
 CC family ligand 6), that binds to the Elk, Nuk/Cek5, Hek2/sek4, Htk
 CC and Sek1 receptors on cells. It was isolated from a human brain
 CC (frontal cortex) library in lambda ZAPII and has been deposited as
 CC pBluescript SK-Efl-6 (ATCC 97319). The isolated DNA can be used to
 CC produce recombinant Efl-6 or soluble truncated Efl-6 polypeptides
 CC in host cells for use in supporting neuronal and other Eph
 CC receptor-bearing cell populations. It can also be used to create
 CC knockout cells, tissues or animals, and in gene therapy.
 SQ Sequence 1860 BP; 282 A; 605 C; 558 G; 409 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x T69808/rev ..

Align seg 1/1 to reverse of: T69808 from: 1 to: 1860

101 LeuGlyGlyGlyArgGlyProSer 108
 ||||||||||||||||||||
 872 CTGGAGGGGGCAGGGGCGCTTCA 849

seq_name: N_Geneseq_36:V06354

seq_documentation_block:
 ID V06354 standard; DNA; 1877 BP.
 AC V06354;
 DE 30-APR-1998 (first entry)
 DE AL-2-long (AL-21) protein encoding DNA.
 KW AL-21; AL-2; AL-2-long; human; treatment; neurological disorder; tumour;
 KW rheumatoid arthritis; wound healing; paralysis; angiogenesis; leukaemia;
 KW psoriasis; Alzheimer's disease; epilepsy; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 244..1611
 FT /*tag= a
 FT /*product= "AL-21"
 FT sig_peptide 244..321
 FT /*tag= b
 FT mat_peptide 322..1608
 FT /*tag= c
 PN WO9740153-A1.
 PD 30-OCT-1997.
 PR 17-APR-1997; U06345.
 PR 19-APR-1996; US-635130.
 PA (GETH) GENENTECH INC.
 PI Caras IW;
 DR WPI: 97-535837/49.
 DR P-PSDB; W33698.
 FT Human AL-2 neurotrophic factor and related DNA - used to develop

PT products for, e.g. treating neurologic disorders, angiogenesis
 PT disorders, tumours or rheumatoid arthritis or for wound healing
 PS Claim 3: Fig 1A-C; 8pp; English.
 CC This DNA encodes a AL-2-long (AL-21) protein. AL-2 is a novel Eph-related
 CC tyrosine kinase receptor ligand. AL-2 can be administered to patients in
 CC whom the nervous system has been damaged by trauma, surgery, stroke,
 CC ischaemia, infection, metabolic disease, nutritional deficiency
 CC malignancy, or toxic agents, to promote the survival or growth of
 CC neurons. They can be used to treat motoneuron disorders such as
 CC amyotrophic lateral sclerosis (Lou Gehrig's disease), Bell's palsy, and
 CC various conditions involving spinal muscular atrophy, or paralysis. AL-2
 CC can be used to treat human neurodegenerative disorders, such as
 CC Alzheimer's disease, Parkinson's disease, epilepsy demyelinating
 CC diseases such as multiple sclerosis, Huntington's chorea, Down's syndrome,
 CC nerve deafness, Menier's disease, and other disorders of the cerebellum.
 CC AL-2 can be used as cognitive enhancer, to enhance learning particularly
 CC in dementias or trauma, since they can promote axonal outgrowth and
 CC synaptic plasticity, particularly of hippocampal neurons that express
 CC AL-2 binding Eph-family receptors and cortical neurons that express
 CC AL-2. AL-2 can also be used for wound healing, i.e. accelerating
 CC neovascularisation of, e.g. burns and ulcers. The encoding nucleic acids
 CC are useful in preparing antibodies that specifically bind to the AL-2
 CC protein. The antibodies and the AL-2 antagonists are useful in diagnosing
 CC and treating various neuronal disorders. AL-2 antagonists can be used
 CC for modulating angiogenesis. They can also be used for the treatment of
 CC tumours, acute myeloid leukaemia (AML), chronic myeloid leukaemia (CML),
 CC myelodysplastic syndrome (MDS), diabetic retinopathy, neovascular
 CC glaucoma, psoriasis and rheumatoid arthritis.
 SQ Sequence 1877 BP; 334 A; 525 C; 615 G; 401 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x V06354/rev ..

Align seg 1/1 to reverse of: V06354 from: 1 to: 1877

101 LeuGlyGlyArgGlyProSer 108
 |||||
 914 CTGGAGGGGCGAGGGGCGCTTCA 891

seq_name: N_Geneseq_36:Q05358

seq_documentation_block:
 ID Q05358 standard; DNA; 1899 BP.
 AC Q05358;
 DT 04-DEC-1990 (first entry)
 DE Glycinin subunit precursor A5A4B3.
 KW Glycinin; glycine; pLGvneol103; ds.
 OS Glycine hispida.
 PN J02156889-A.
 PD 15-JUN-1990.
 PF 08-DEC-1988; 310553.
 PR 08-DEC-1988; JP-310553.
 PA (NOR) NORINSHO.
 PA (KIRI) KIRIN BREWERY KK.
 DR WPI: 90-22848/30.
 PT Recombinant plasmid - obtd. by recombining glycinin gene to
 PT plasmid for plant.
 PS Claim 5; Page 561; 20pp; Japanese.
 CC Glycinin precursor gene may be used with exogenous promoter and
 CC terminator in plasmid pLGvneol103 to transform plants, improving
 CC nutritional value
 SQ Sequence 1899 BP; 583 A; 468 C; 439 G; 408 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x Q05358 ..

Align seg 1/1 to: Q05358 from: 1 to: 1899

327 AspGluAspGluAspGluasp 334
 |||||
 884 GATGAAGACGAGATGATGAAGAT 907

seq_name: N_Geneseq_36:T45793

seq_documentation_block:

ID T45793 standard; cDNA; 1934 BP.
 AC T45793;
 DT 10-MAR-1997 (first entry)
 DE Heliothis ecdysona steroid receptor cDNA clone pSK19R.
 KW Ecdysone steroid receptor; HECK; gene switch; insect resistance;
 KW herbicide resistance; transgenic plant; cancer; gene therapy; ds;
 KW cyclic.
 OS Heliothis virescens.
 PN W09637609-A1.
 PD 28-NOV-1996.
 PF 20-MAY-1996; G01195.
 PR 26-MAY-1995; GB-010759.
 PR 07-JUL-1995; GB-013882.
 PR 24-AUG-1995; GB-017316.
 PR 18-MAR-1996; GB-005656.
 PA (ZENE) ZENECA LTD.
 PI Greenland AJ; Jepson I; Martinez A;
 DR WPI: 97-033992/03.
 PT DNA encoding insect ecdysona steroid receptor - acts as a gene
 PT switch responsive to chemical induction enabling external control of
 PT the gene
 PS Claim 1; Page 41-42; 122pp; English.
 CC A cDNA sequence (T45793) is contained within clone pSK19R (NCIMB
 CC 40743) isolated from a random primed Heliothis virescens 4th and
 CC 5th instar library, and codes for part of the Heliothis ecdysona
 CC steroid receptor (HECR). It was detected using a partial clone
 CC (see also T45797) contg. sequences matching the DNA binding domain
 CC of the Drosophila ecdysona steroid receptor. A probe contg. the
 CC 5' end of pSK19R was used to rescreen the library, yielding
 CC plasmid pSK16.1 (T45794). 5'RACE (see also T45803-07) was used
 CC to obtain the full open reading frame (T45795) coding for HECR
 CC (W06533). HECR clones, esp. sequences coding for the ligand
 CC binding domain, can be used as gene switches, allowing inducible
 CC control of foreign genes in e.g. transgenic plants or mammals.
 SQ Sequence 1934 BP; 481 A; 547 C; 532 G; 374 T;

alignment_scores:

Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x T45793 ..

Align seg 1/1 to: T45793 from: 1 to: 1934

329 AspGluAspGluAspSerAsp 336
 |||||
 1094 GACGAGGACGACGACGATCGGAT 1117

seq_name: N_Geneseq_36:Q21400

seq_documentation_block:

ID Q21400 standard; DNA; 2000 BP.
 AC Q21400;
 DT 03-JUN-1992 (first entry)
 DE Vibriolysin gene sequence.
 KW Protease; wound treatment; necrotic tissue; collagen; elastin;

KW fibrin; wound debridement; vascularisation; ss.
 OS Vibrio proteolyticus.
 FH Key Location/Qualifiers
 FT CDS 61..1890
 FT /tag= a
 FT /product= vibriolysin
 FT
 EP-472011-A.
 PN 26-FEB-1992.
 PD 29-JUL-1991; 112732.
 PF 15-AUG-1990; US-567884.
 PR 13-MAR-1991; US-670612.
 PA (GRAC) GRACE W R & CO-CONN.
 PI Fortney DZ, Durham DR;
 DR WPI: 92-066263/09.
 DR P-PSDB; R21414.
 PT Protease compsn. for wound treatment - contains protease from
 PT Vibrio, esp. V proteolyticus, which hydrolyses components of
 PT necrotic tissue but not native tissue
 PS Claim 3; Fig 1; 23pp; English.
 CC The vibriolysin protease can be produced by aerobic fermentation of
 CC Vibrio species in nutrient medium. The DNA sequence was obt'd. by
 CC cloning the Vibrio protease. A gene library was prepd. using Vibrio
 CC chromosomal DNA. The DNA underwent partial digestion with Sau3A
 CC and ligation into E. coli cosmid vector pHC79. The recombinant
 CC vectors were packaged into bacteriophage lambda and used to
 CC transform E. coli strain HB101. Clones contg. vibriolysin were
 CC found to produce a zone of clearing on milk agar plates due to
 CC proteolytic hydrolysis of the casein component of milk. The
 CC living tissue is capable of digesting necrotic tissue while viable
 CC living tissue is not injured. It causes wound debridement and
 CC stimulates vascularisation and healing of traumatised tissue. The
 CC protease can be used in management of full and partial thickness
 CC wounds, burn wounds, debridement of ulcerative lesions, principally
 CC pressure (decubitus) ulcers and varicose stasis and trophic ulcers,
 CC prepn. of skin graft sites and general surgical wounds such as
 CC amputation, incisional, traumatic and pyogenic wounds and for the
 CC treatment of vaginitis, cervicitis, circumcisions, epistomy, cyst
 CC wounds, carbuncles, sunburn, frostbite and cataract scar tissue.
 CC Sequence 2000 BP; 537 A; 467 C; 502 G; 494 T;
 SQ

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x Q21400/rev ..
 Align seg 1/1 to reverse of: Q21400 from: 1 to: 2000

639 TyrIleAlaArgArgArgSer 646
 |||||
 1596 TACATCGCCAGACGAGGAGGAGT 1573

seq_name: N_Geneseq_36:V84700

seq_documentation_block:
 ID V84700 standard; DNA; 2000 BP.
 AC V84700;
 DT 29-MAR-1999 (first entry)
 DE Neutral protease vibriolysin DNA.
 KW bone fracture; surgical abrasion; bed sore; ulcer; tendonitis;
 KW bursitis; vaginitis; cervicitis; circumcisions; epistomy;
 KW pilonidal cyst wart; carbuncle; sunburn; frostbite; therapy; ss.
 OS Vibrio proteolyticus strain ATCC 53559.
 FH Key Location/Qualifiers
 FT CDS 61..1890
 FT /tag= a
 FT WO9855604-A1.
 PN 10-DEC-1998.
 PD 01-JUN-1998; U10698.

PR 02-JUN-1997: US-867331.
 PA (GRAC) GRACE & CO-CONN W R.
 PI Durham DR, Fortney DZ, Yang K;
 DR WPI: 99-070263/06.
 DR P-PSDB; W86187.
 PT Hydrophilic composition containing enzyme, especially protease, and
 PT glyceryl cocoate - stable for months at room temperature.
 PT specifically used to debride wounds and promote their healing
 PS Claim 7; Page 20-25; 31pp; English.
 CC This nucleotide sequence codes for vibriolysin (see W86187), a new
 CC extracellular neutral protease of Vibrio proteolyticus ATCC 53559.
 CC Vibriolysin can be obtained by fermentation of V. proteolyticus,
 CC or by cultivation of recombinant host cells, collecting the enzyme
 CC from the culture broth. A claimed hydrophilic pharmaceutical
 CC composition comprises an extracellular neutral protease produced by
 CC Vibrio and enough glyceryl cocoate to maintain enzymatic activity
 CC at over 80% for at least 100 days at room temperature. The
 CC composition is used to debride wounds and to promote wound healing,
 CC more generally to remove any necrotic and/or non-viable tissue.
 CC Particular applications are treatment of burns, bone fractures,
 CC surgical abrasions, bed sores, slowly healing ulcers, tendonitis,
 CC bursitis, vaginitis, cervicitis, circumcisions, epistomy,
 CC pilonidal cyst warts, carbuncles, sunburn and frostbite (claimed).
 CC Vibriolysin is able to hydrolyse components of eschar (including
 CC denatured collagen, elastin and fibrin) but does not damage native
 CC tissue.
 SQ Sequence 2000 BP; 537 A; 467 C; 502 G; 494 T;
 alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x V84700/rev ..
 Align seg 1/1 to reverse of: V84700 from: 1 to: 2000

639 TyrIleAlaArgArgArgSer 646
 |||||
 1596 TACATCGCCAGACGAGGAGGAGT 1573

seq_name: N_Geneseq_36:T84986

seq_documentation_block:
 ID T84986 standard; DNA; 2361 BP.
 AC T84986;
 DT 05-DEC-1997 (first entry)
 DE Burkholderia cepacia insertion sequence hybrid IS402/1356.
 KW Hybrid; insertion; element; sequence; IS402; IS1356; detection;
 KW polymerase chain reaction; PCR; enzyme linked immunosorbant assay;
 KW ELISA; virulent strain; cystic fibrosis; ss.
 OS Burkholderia cepacia.
 OS Pseudomonas cepacia.
 FH Key Location/Qualifiers
 FT CDS 243..1509
 FT /tag= a
 FT WO9707237-A1.
 PN 27-FEB-1997.
 PD 16-AUG-1996; CA0550.
 PR 17-AUG-1995; US-002398.
 PA (CNDG) CANADA MIN HEALTH.
 PI Johnson WM, Rozee KR, Tyler SD;
 DR WPI: 97-165317/15.
 DR P-PSDB; W27464.
 PT New insertion element from Burkholderia cepacia - used for
 PT identifying virulent isolates of B. cepacia and other organisms.
 PS Claim 2; Pages 27-28; 53pp; English.
 CC The present insertion element is a hybrid of the known Burkholderia
 CC cepacia strain ET12, formerly known as Pseudomonas cepacia,
 CC insertion sequences IS402 and IS1356. Detection of the insertion
 CC element in a sample, e.g. by polymerase chain reaction (PCR) or

CC enzyme linked immunosorbant assay (ELISA), indicates that a
 CC virulent isolate of B. cepacia or another pathogenic organism is
 CC present.

CC PCR was carried out using primers for the B. cepacia elements
 CC IS1356, IS402, IS406, IS407 and IS408. The primers targeting
 CC IS402, in addition to detecting the expected IS element, also
 CC primed an amplicon of 650 bp in some isolates. This was restricted
 CC to isolates known to be highly transmissible in cystic fibrosis
 CC patients. Sequence data revealed that the 650 bp amplicon consisted
 CC initially of the IS402 sequence, but this was interrupted after 154
 CC bp and succeeded by that of IS1356.
 SQ Sequence 2361 BP; 473 A; 720 C; 737 G; 431 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x T84986/rev ..

Align seg 1/1 to reverse of: T84986 from: 1 to: 2361

500 AlailePheSerLysAlaAlaAla 507

|||||
 2147 GCATATTTCGAAGCCGCTGCC 2124

seq_name: N_Geneseq_36:N90726

seq_documentation_block:

ID N90726 standard; DNA; 2377 BP.

AC N90726;

DE 19-JUN-1990 (first entry)

DF DNA sequence of extracellular neutral protease (vibriolysin) gene from

DE Vibrio proteolyticus ATCC 53559

KW Extracellular neutral protease gene; vibriolysin gene;

OS Vibrio proteolyticus; ATCC 53559.

FS Vibrio proteolyticus.

FH Key Location/Qualifiers

FT cds 3..35

FT /*tag= a

FT /*note="p93568"

FT cds 39..68

FT /*tag= b

FT /*note="p93569"

FT cds 72..86

FT /*tag= c

FT /*note="p93570"

FT cds 90..104

FT /*tag= d

FT /*note="p93571"

FT cds 108..116

FT /*tag= e

FT cds 120..155

FT /*tag= f

FT /*note="p93573"

FT cds 159..200

FT /*tag= g

FT /*note="p93575"

FT cds 204..215

FT /*tag= h

FT /*note="p93576"

FT cds 219..244

FT /*tag= i

FT /*note="p93577"

FT cds 248..2078

FT /*tag= j

FT /*note="p93578"

FT cds 2082..2111

FT /*tag= k

FT /*note="p93579"

FT cds 2115..2144

FT /*tag= l

FT /*note="p93580"

FT cds 2148..2150

FT /*tag= m

FT cds 2154..2240

FT /*tag= n

FT /*note="p93582"

FT cds 2244..2309

FT /*tag= o

FT /*note="p93583"

FT cds 2313..2317

FT /*tag= p

FT cds 2321..2324

FT /*tag= q

FT cds 2328..2377

FT /*tag= r

FT /*note="p93586"

PN EP-309879-A.

PD 05-APR-1989.

PF 21-AUG-1988; 115439.

PR 19-NOV-1987; US-123038, US-103983.

PA (GRAC) Grace WR Co.

PI Deutch AH, David VA;

DR WPI: 89-101218/14.

DR P-PSDB; P93568, P93569, P93570, P93571, P93572, P93573, P93575, P93576, P93577,

DR P93578, P93579, P93580, P93582, P93583, P93586.

PT Cloning and expression of neutral protease genes -

PT using DNA encoding enzymes of Vibrio proteolyticus or Bacillus

PT species in gram-negative microorganisms

PS Figure 1; 16pp; English.

CC Figure 1 gives the translation of the entire sequence, hence the large

CC number of CDS, which are separated by stop codons. Recombinant DNA

CC contg. a Vibrio nuclear protease enzyme (NPE) gene is claimed as are

CC gram-negative microorganisms contg. DNA coding from more than one NPE of

CC Vibrio proteolyticus. The microorganisms are pref. E.coli or Serratia

CC species, esp. E.coli ATCC 67499 or 67501. The gram-negative

CC microorganisms synthesise active protease enzyme. The genes can be

CC manipulated for the overprodn. of NPE.

SQ Sequence 2377 BP; 646 A; 527 C; 585 G; 618 T; 1 Others;

alignment_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x N90726/rev ..

Align seg 1/1 to reverse of: N90726 from: 1 to: 2377

639 TyrileAlaArgArgArgSer 646

|||||

1787 TACATCGCCAGCGAGGAGGAGT 1764

seq_name: N_Geneseq_36:Q68538

seq_documentation_block:

ID Q68538 standard; DNA; 2377 BP.

AC Q68538;

DT 20-FEB-1995 (first entry)

DE Extracellular neutral protease (vibriolysin) gene.

KW Extracellular; neutral protease; vibriolysin; V. proteolyticus;

KW recombinant; gram-negative; microorganisms; E. coli; bacteria;

KW secretion; lysis; ds.

OS Vibrio proteolyticus ATCC 53559.

FH Key Location/Qualifiers

FT cds 249..2081

FT /*tag= a

FT /*note="large open reading frame - encodes neutral

FT protease"

PN EP-605073-A.

PD 06-JUL-1994.

PF 21-SEP-1988; 115439.
 PR 01-OCT-1987; US-103983.
 PR 19-NOV-1987; US-123038.
 PA (GRAC) GRACE & CO-CONN W R.
 PI David VA, Deutch AH;
 DR WPI; 94-210320/26.
 DR P-PSDB; R59765.
 PT Recombinant DNA contg neutral protease genes - obtd from Vibrio
 PT species or Bacillus stearothermophilus, used for large scale
 PT prodn of the enzyme.
 PS Claim 5; Fig 1; 15pp; English.
 CC This sequence encodes the extracellular neutral protease (vibriolysin)
 CC from V. proteolyticus ATCC 53559. This sequence may be used in the
 CC production of recombinant gram-negative microorganisms, such as E. coli,
 CC which secretes the heterologous protease. The enzyme may also be
 CC liberated by lysis of the microbial cell. Recombinant E. coli producing
 CC the protease may be used for the large scale production of active,
 CC functional, neutral protease enzymes for industrial use.
 SQ Sequence 2377 BP; 646 A; 528 C; 586 G; 617 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x Q68538/rev ..
 Align seg 1/1 to reverse of: Q68538 from: 1 to: 2377

639 TyrIleAlaArgArgArgSer 646
 |||||
 1787 TACATCCGACGAGGAGGAGT 1764

seq_name: N_Geneseq_36:V06355

seq_documentation_block:

ID V06355 standard; DNA; 2380 BP.
 AC V06355;
 DT 30-APR-1998 (first entry)
 DE AL-2-short (AL-2s) protein encoding DNA.
 KW AL-2; AL-2s; AL-2-short; human; treatment; neurological disorder;
 KW rheumatoid arthritis; wound healing; paralysis; angioedema; leukaemia;
 KW tumour; psoriasis; Alzheimer's disease; epilepsy; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 244..1266
 FT /*tag= a
 FT /product= "AL-2s"
 FT sig_peptide 244..321
 FT /*tag= b
 FT met_peptide 322..1263
 FT /*tag= c
 PN W09740153-AL.
 PD 30-OCT-1997.
 PF 17-APR-1997; U06345.
 PR 19-APR-1996; US-635130.
 PA (GETH) GENENTECH INC.
 PI Caras IW;
 DR WPI; 97-535837/49.
 DR P-PSDB; W33699.
 PT Human AL-2 neurotrophic factor and related DNA - used to develop
 PT products for, e.g. treating neurologic disorders, angioedema
 PT disorders, tumours or rheumatoid arthritis or for wound healing
 PS Claim 4; Fig 2A-B; 86pp; English.
 CC This DNA encodes a AL-2-short (AL-2s) protein. AL-2 is an Eph-related
 CC tyrosine kinase receptor ligand. AL-2 can be administered to patients in
 CC whom the nervous system has been damaged by trauma, surgery, stroke,
 CC ischaemia, infection, metabolic disease, nutritional deficiency,
 CC malignancy, or toxic agents, to promote the survival or growth of
 CC neurons. They can be used to treat motoneuron disorders such as
 CC amyotrophic lateral sclerosis (Lou Gehrig's disease), Bell's palsy, and

CC various conditions involving spinal muscular atrophy, or paralysis. AL-2
 CC can be used to treat human neurodegenerative disorders, such as
 CC Alzheimer's disease, Parkinson's disease, epilepsy, demyelinating
 CC diseases such as multiple sclerosis, Huntingtons chorea, Down's syndrome,
 CC nerve deafness, Menier's disease, and other disorders of the cerebellum.
 CC AL-2 can be used as cognitive enhancer, to enhance learning particularly
 CC in dementias or trauma, since they can promote axonal outgrowth and
 CC synaptic plasticity, particularly of hippocampal neurons that express
 CC AL-2 binding Eph-family receptors and cortical neurons that express
 CC AL-2. AL-2 can also be used for wound healing, i.e. accelerating
 CC neovascularisation of, e.g. burns and ulcers. The encoding nucleic acids
 CC are useful in preparing antibodies that specifically bind to the AL-2
 CC protein. The antibodies and the AL-2 antagonists are useful in diagnosing
 CC and treating various neuronal disorders. AL-2 antagonists can be used
 CC for modulating angiogenesis. They can also be used for the treatment of
 CC tumours, acute myeloid leukaemia (AML), chronic myeloid leukaemia (CML),
 CC myelodysplastic syndrome (MDS), diabetic myeloid leukaemia (CML),
 CC glaucoma, psoriasis and rheumatoid arthritis.
 SQ Sequence 2380 BP; 404 A; 722 C; 736 G; 516 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x V06355/rev ..
 Align seg 1/1 to reverse of: V06355 from: 1 to: 2380

101 LeuGlyGlyClyArgGlyProSer 108
 |||||
 914 CTGGGAGGGGGCAGGGGCTTCA 891

seq_name: N_Geneseq_36:T45794

seq_documentation_block:

ID T45794 standard; cDNA; 2464 BP.
 AC T45794;
 DT 10-MAR-1997 (first entry)
 DE Heliothis eclypsone steroid receptor cDNA clone pSK16.1.
 KW Eclypsone steroid receptor; HEGR; gene switch; insect resistance;
 KW herbicide resistance; transgenic plant; cancer; gene therapy; ds;
 KW cyclic.
 OS Heliothis virescens.
 FH Key Location/Qualifiers
 FT primer_bind complement (1..24)
 FT /*tag= a
 FT /note= "primer 16PCR1 binding site"
 FT misc_difference 2241
 FT /*tag= b
 FT /note= "base 2241 is given as n in the
 FT specification"
 PN W09637609-AL.
 PD 28-NOV-1996.
 PF 20-MAY-1996; G01195.
 PR 26-MAY-1995; GB-010759.
 PR 07-JUL-1995; GB-013882.
 PR 24-AUG-1995; GB-017316.
 PR 18-MAR-1996; GB-005656.
 PA (ZENE) ZENECA LTD.
 PI Greenland AJ, Jepsen I, Martinez A;
 DR WPI; 97-033992/03.
 PT DNA encoding insect eclypsone steroid receptor - acts as a gene
 PT switch responsive to chemical induction enabling external control of
 PT the gene
 PS Claim 2; Page 42-43; 122pp; English.
 CC A cDNA sequence (T45794) is contained within clone pSK16.1 isolated
 CC from a random primed Heliothis virescens 4th and 5th instar library
 CC and codes for part of the Heliothis eclypsone steroid receptor
 CC (HEGR). It was detected using a probe contg. the 5' end of clone
 CC pSK19R (see also T45793). 5'RACE (see also T45803-07) was used to

CC obtain the full open reading frame (T45795) coding for HECr
 CC (W06533). HECr clones, esp. sequences coding for the ligand
 CC binding domain, can be used as gene switches, allowing inducible
 CC control of foreign genes in e.g. transgenic plants or mammals.
 SQ Sequence 2464 BP; 609 A; 633 C; 681 G; 540 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x T45794 ..

Align seg 1/1 to: T45794 from: 1 to: 2464

329 AspGluAspGluAspSerAsp 336
 |||||
 973 GACGAGGACGACGAGACTCGGAT 996

seq_name: N_Geneseq_36:T45795

seq_documentation_block:

ID T45795 standard; cDNA; 2745 BP.
 AC T45795;
 DT 10-MAR-1997 (first entry)
 DE Heliothis ecdyson steroid receptor cDNA.
 KW Ecdysone steroid receptor; HECr; gene switch; insect resistance;
 KW herbicide resistance; transgenic plant; cancer; gene therapy; ds;
 KW cyclic.
 OS Heliothis virescens.
 FH Key Location/Qualifiers
 FT cds 225..1955
 FT /*tag= a
 FT /transl_except= 300..302; aa:Gly
 FT /note= "CCA codes for proline, as shown in Fig 4"
 FT misc_difference 2522
 FT /*tag= b
 FT /note= "base 2522 is given as n in the
 specification"

WO9637609-A1.
 28-NOV-1996.
 PD 20-MAY-1996; G01195
 PF 26-MAY-1995; GB-010759.
 PR 07-JUL-1995; GB-013882.
 PR 24-AUG-1995; GB-017316.
 PR 18-MAR-1996; GB-005656.
 PA (ZENE) ZENECA LTD.
 PI Greenland AJ, Jepson I, Martinez A;
 DR WPI; 97-033992/03.
 DR P-PSDB; W06533.
 PT DNA encoding insect ecdyson steroid receptor - acts as a gene
 PT switch responsive to chemical induction enabling external control of
 PT the gene
 PS Claim 3; Page 44-45; 122pp; English.

CC A cDNA sequence (T45795) includes an open reading frame coding
 CC for Heliothis virescens ecdyson steroid receptor (HECr) (W06533).
 CC It was deduced from 5'RACE products (see also 45803-07) fused to
 CC the sequence of clone pSKI6.1 (see also T45794), obtd. from a
 CC H. virescens 4th and 5th instar cDNA library. HECr clones, esp.
 CC sequences coding for the ligand binding domain, are useful as gene
 CC switches, allowing external control of foreign genes to which they
 CC are linked e.g. to confer herbicide resistance or insect tolerance
 CC to transgenic plants, and to allow the timing of expression of a
 CC therapeutic gene to be controlled in mammals.
 SQ Sequence 2745 BP; 671 A; 694 C; 767 G; 612 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x T45795 ..

Align seg 1/1 to: T45795 from: 1 to: 2745

329 AspGluAspGluAspSerAsp 336
 |||||
 1254 GACGAGGACGACGAGACTCGGAT 1277

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

OM of: US-09-332-522B-2 to: Issued_Patents_NA.* out_format : pfs
 Date: Sep 27, 2000 2:47 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL-frame+ p2n.model -DEV-xlp
 -O-cg2_1/USPO.spool/US09332522/runat-27092000_091221_6548/app_query.fasta_1.1355
 -DB-Issued_Patents_NA -QPMF-fastap -SUFFIX-oli.xni -GAPOP-4.500
 -GAPEXT-0.050 -MINMATCH-0.100 -LOOPCL-0.000 -LOOPEXT-0.000
 -GAPOP-4.500 -GAPEXT-0.050 -GAPOP-60.000 -XGAPEXT-60.000
 -GAPOP-6.000 -GAPEXT-7.000 -GAPOP-60.000 -XGAPEXT-60.000
 -DELOP-6.000 -DELEXT-7.000 -START-1 -MATRIX-oligo
 -TRANS-human40.cdi -LIST-45 -DLOCALIGN-200 -THR SCORE-quality
 -THR_MIN-1 -ALIGN-45 -MODE-LOCAL -OUTFMT-pfs -NORM-ext -MINLEN-0
 -MAXLEN-200000000 -USER-US09332522 -CGN1_1_92 -NCFU-6 -ICPU-3
 -LONGLOG -NO_XLPXY -WAIT -THREADS-1

Search information block:
 Query: US-09-332-522B-2
 Query length: 1113
 Database: Issued_Patents_NA.*
 Database sequences: 243080
 Database length: 6877915
 Search time (sec): 77.950000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP-XGAPOP-60.000
 WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT-XGAPEXT-60.000

Sequence	Strd Orig	ZScore	PScore	Len	Documentation
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-28	54	1	9.00	145.45	1.30
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-29	54	1	9.00	145.45	1.30
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-30	54	1	9.00	145.45	1.30
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-31	789	1	9.00	126.01	15.76
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-32	961	1	9.00	124.58	18.93
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-33	961	1	9.00	124.58	18.93
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-34	1282	1	9.00	122.49	24.74
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-35	1543	1	9.00	121.15	29.39
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-36	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-37	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-38	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-39	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-40	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-41	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-42	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-43	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-44	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-45	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-46	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-47	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-48	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-49	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-50	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-51	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-52	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-53	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-54	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-55	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-56	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-57	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-58	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-59	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-60	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-61	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-62	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-63	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-64	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-65	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-66	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-67	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-68	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-69	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-70	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-71	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-72	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-73	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-74	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-75	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-76	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-77	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-78	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-79	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-80	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-81	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-82	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-83	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-84	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-85	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-86	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-87	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-88	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-89	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-90	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-91	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-92	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-93	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-94	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-95	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-96	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-97	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-98	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-99	1744	1	9.00	120.26	32.94
/cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-100	1744	1	9.00	120.26	32.94

/cg2_6/ptodata/2/ina/5A_COMB.seq:US-07-807-043B-5 + 8.00 96.26 715.47 4698
 /cg2_6/ptodata/2/ina/5A_COMB.seq:US-08-299-849B-5 + 8.00 96.26 715.47 4698
 /cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-142-368A-5 + 8.00 96.26 715.47 4698
 /cg2_6/ptodata/2/ina/6_COMB.seq:US-08-967-737-5 + 8.00 96.26 715.47 4698
 /cg2_6/ptodata/2/ina/5A_COMB.seq:US-08-131-365B-53 + 8.00 95.53 785.86 5197
 /cg2_6/ptodata/2/ina/5C_COMB.seq:US-08-668-123-53 + 8.00 95.53 785.86 5197

seq_name: /cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-28

seq_documentation_block:
 ; Sequence 28, Application US/08649991
 ; Patent No. 5919462

GENERAL INFORMATION:

APPLICANT: Naita, Remy

APPLICANT: Roques, Pierre

TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE

TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR

TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF

TITLE OF INVENTION: MATERNOPREVENTION TRANSMISSION OF HIV-1

NUMBER OF SEQUENCES: 130

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP

STREET: 1800 M Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20036-5869

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/649,991

FILING DATE: 17-MAY-1996

CLASSIFICATION: 435

PRIOR APPLICATION NUMBER: 435

APPLICATION DATA: FR 9505914

FILING DATE: 18-MAY-1995

ATTORNEY/AGENT INFORMATION:

NAME: Adler, Reid G.

REGISTRATION NUMBER: 30,988

REFERENCE/DOCKET NUMBER: ORES-5003

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-467-7000

TELEFAX: 202-467-7176

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 54 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-649-991-28

alignment_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x US-08-649-991-28/rev

Align seq 1/1 to reverse of: US-08-649-991-28 from: 1 to: 54

972 LeuphSerThrSerArgAlaSer 980

|||||

29 CTCTTTCCACTTCTCTAGACTTCT 3

seq_name: /cg2_6/ptodata/2/ina/5D_COMB.seq:US-08-649-991-29

seq_documentation_block:

THIS PAGE BLANK (USPTO)

OM of: US-09-332-522B-2 to: GenEmbl.* out_format : pfs

Date: Sep 27, 2000 2:44 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame-p2n.model -DEV=xlp  
-o=/gen2_1/USPTO_spool/US09332522/runat_27092000_091221_6535/app_query.fasta_1.1355  
-DB=GenEmbl -QFMT=fastap -SUFFIX=oli.rge -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=60.000 -GAPEXT=60.000  
-GAPOP=6.000 -GAPEXT=7.000 -GAPOP=60.000 -GAPEXT=60.000  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=quality  
-THR_MIN=1 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM-ext -MINLEN=0  
-MAXLEN=200000000 -USER=US09332522 -CGN1.1.6998 -NCFU=6  
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-332-522B-2

Query length: 1113

Database: GenEmbl.*

Database sequences: 972840

Database length: 892348106

Search time (sec): 3472.210000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000

WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

seq	stnd	Orig	EScore	Len	Documentation
gb_in1:CEY47D3B	-	176.00	3294.32	2.1e-175	95968 AL031635 Caenorhabditis eleg
gb_htg1:CEY66A7	-	176.00	3289.06	4.2e-175	197735 AL022282 Caenorhabditis eleg
gb_htg1:CEY47D3	-	176.00	3285.17	6.9e-175	197735 298865 Caenorhabditis eleg
gb_htg1:CEH10N23	-	153.00	2861.69	2.7e-151	36780 294157 Caenorhabditis elegans
gb_htg1:CEH10N23	-	126.00	2345.64	1.5e-122	36780 294157 Caenorhabditis elegans
gb_htg1:CEY48A6	-	120.00	2215.78	2.5e-115	296699 292854 Caenorhabditis eleg
gb_pr3:AF039153	+	10.00	157.01	7.32	AF039153 Homo sapiens chromosom
gb_pr3:AF039153	+	10.00	156.67	1.25	767 J02512 Bacteriophage T4 genes 6
gb_p11:AB000209	+	10.00	154.33	1.69	1059 AB000209 Nannochloropsis ocula
gb_p11:AB000209	+	10.00	136.86	16.25	12039 AC014381 Drosophila melanog
gb_p11:AF117653	+	10.00	135.83	18.07	13479 AF117653 Homo sapiens double
gb_pr4:AF118101	+	10.00	117.45	190.93	168900 AF118101 Coliphage T4, compl
gb_htg21:AC011080	+	10.00	116.99	202.52	179921 AC011080 Homo sapiens clone
gb_in2:AE003429	-	10.00	113.33	323.61	297405 AE003429 Drosophila melanog
gb_pat:AS7962	-	9.00	156.85	1.22	54 AS7962 Sequence 28 from Patent B
gb_pat:AS7962	-	9.00	136.85	1.22	54 AS7962 Sequence 29 from Patent B
gb_ro:MUSIGHV2	+	9.00	146.51	4.59	224 M17576 Mouse Ig heavy-chain var
gb_v11:HIV1GNC	+	9.00	142.53	7.65	387 279576 Human immunodeficiency v
gb_v11:HIV1DUBC	-	9.00	142.37	7.81	396 279575 Human immunodeficiency v
gb_v11:HIV1MPAC	-	9.00	142.37	7.81	396 279574 Human immunodeficiency v
gb_v11:HIV1MPAM	-	9.00	142.37	7.81	396 279573 Human immunodeficiency v
gb_pr3:AF039150	+	9.00	137.88	13.89	734 AF039150 Homo sapiens chromosom
gb_pr3:AF039151	+	9.00	137.88	13.89	734 AF039151 Homo sapiens chromosom
gb_v11:AF039152	+	9.00	137.87	13.91	735 AF039152 Homo sapiens chromosom
gb_pr3:AF039154	+	9.00	137.87	13.91	735 AF039154 Homo sapiens chromosom
gb_pr3:AF039155	+	9.00	137.87	13.91	735 AF039155 Homo sapiens chromosom
gb_om:BOVVEGFB	+	9.00	137.24	15.09	802 M33750 Bovine vascular endothel
gb_om:BOVVEGFA	+	9.00	136.13	17.39	934 M18336 Bovine vascular endothel
gb_v11:AB021155	+	9.00	135.82	18.11	975 AB021155 Unidentified baculovir
gb_pat:AS2895	+	9.00	133.82	23.37	1282 AB2895 R.prolifus throbmin inh
gb_v11:AB021155	+	9.00	133.82	23.37	1282 121698 Sequence 16 from patent
gb_v11:AB021155	+	9.00	133.16	25.46	1405 AJ012282 Myxoma virus mt3, t3c
gb_v11:HIV1B2A17	-	9.00	132.87	26.42	1462 L11778 Human immunodeficiency
gb_om:BOVVEGFA	+	9.00	132.51	27.66	1636 M32976 Bovine heparin-binding
gb_om:AB025793	+	9.00	132.14	29.00	1516 AB025793 Octopus vulgaris Ove
gb_pat:114451	+	9.00	131.59	31.14	1744 I14451 Sequence 25 from patent
gb_pat:127294	+	9.00	131.59	31.14	1744 I27294 Sequence 25 from patent
gb_in1:EBU67116	+	9.00	131.59	31.14	1744 U67116 Emerita brunetti small
gb_in1:CFTRG	+	9.00	131.48	31.59	1771 Z12618 C.fasciculata gene enco

gb_ro:MUSIGHV11 + 9.00 131.27 32.45 1823 | J00503 Mouse Ig germline H-
gb_in1:TRFTRPRED8 + 9.00 131.27 32.45 1823 | M73324 Crithidia fascicula-
gb_in1:TRFTRPREDC - 9.00 131.27 32.45 1823 | M73325 Crithidia fascicula-
gb_ov:AF060951 + 9.00 130.35 36.50 2068 | AF060951 Gallus gallus home
gb_ro:MMIGVH23 + 9.00 130.18 37.28 2115 | X06856 Mouse immunoglobulin
gb_ro:MUSIGHV1 + 9.00 130.18 37.31 2117 | M17571 Mouse Ig heavy-chain

seq_name: gb_in1:CEY47D3B

seq_documentation_block:

LOCUS CEY47D3B 95968 bp DNA INV 14-DEC-1999
DEFINITION Caenorhabditis elegans cosmid Y47D3B, complete sequence.
ACCESSION AL031635 298865
VERSION AL031635.1 GI:3646936
HTG.

SOURCE
Caenorhabditis elegans.

ORGANISM
Caenorhabditis elegans.

REFERENCE
1 (bases 1 to 95968)

AUTHORS
none.

TITLE
Genome sequence of the nematode C. elegans: a platform for

JOURNAL
Investigating biology. The C. elegans Sequencing Consortium

MEDLINE
Science 282 (5396), 2012-2018 (1998)

REMARK
90069613

REFERENCE
The C.elegans Sequencing Consortium.

AUTHORS
Erratum: [[published errata appear in Science 1999 Jan

TITLE
1:283(5398):35 and 1999 Mar 26:283(5410):2103]]

JOURNAL
2 (bases 1 to 95968)

MEDLINE
Matthews, L.

REMARK
Direct Submission

AUTHORS
Submitted (22-SEP-1998) Nematode Sequencing Project, Sanger Centre,

TITLE
Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,

JOURNAL
Washington University, St. Louis, MO 63110, USA. E-mail:

COMMENT
jes@sanger.ac.uk or tw@nematoe.wustl.edu

coding sequences below are predicted from computer analysis, using

predictions from Genefinder (P. Green, U. Washington), and other

available information.

For a graphical representation of this sequence and its analysis

see:

http://webace.sanger.ac.uk/cgi-

bin/display?db=wormbase&class=Sequence&object=Y47D3B

Current sequence finishing criteria for the C. elegans genome

sequencing consortium are that all bases are either sequenced

unambiguously on both strands, or on a single strand with both a

dye primer and dye terminator reaction, from distinct subclones.

Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of

the specified clone. It may be shorter because we only sequence

overlapping sections once, or longer because we arrange for a small

overlap between neighbouring submissions.

IMPORTANT: This sequence is not the entire insert of clone Y47D3B.

It may be shorter because we only sequence overlapping sections

once, or longer because we arrange for a small overlap between

neighbouring submissions.

The true left end of clone H10N23 is at 61856 in this sequence. The

true right end of clone T2806 is at 104 in this sequence. The true

right end of clone Y47D3 is at 95968 in this sequence. The start of

this sequence (1..104) overlaps with the end of sequence 281134.

The end of this sequence (95863..95968) overlaps with the start of

sequence AL032622.

Location/Qualifiers

1..95968

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="III"

/clone="Y47D3B"

7188..12736

/gene="Y47D3B.1"

join(7188..7235,7287..7436,7488..7605,7654..7713,

8884..9277,10132..10318,11510..11678,11731..11860,

12379..12481,12657..12736)

/gene="Y47D3B.1"

FEATURES

source

gene

CDS

```

/codon_start=1
/protein_id="CAA21039.1"
/db_xref="GI:3881005"
/translation="MWAKNNRYDIFLRKIDINTROIILGSFWRDEREQNDGRFAVIFH
LGDRIEHLPLCFISIOANNRGLITATVORTHRKASSDICSWAGHAEGLDVLH
KEFISTPNVEDAMLTPELPREPHDMVVCAPWYITDWEILVTELTAMGA
TKIVVQSNADAMRILQIYERKGLVIRWPKWPMDSNDPNGLVLSRGIESHYN
CLFVPSWSDIVFSDIDFLPDSTISPDNLQILKNIFAHPHAGSLLFEHRDV
FVPPNRQGSALFNFEFLSHSKNKNMNCNVRKTRVVVNAKSRVNNHTEGIR
FGYQTRVPCQAHFYHLRHSNHTVPSTPINMSPADMLNKQWOTVEGTGMKDEV
LNKNTVEFDRCMGAINDEHWTMKVSRCLTHVCFSLRRLREVDCVAVASEFEYR
SGKVRIOAEVNCSPVPTSGNHYPAP"
complement(12895..17349)
/gene="Y47D3B.2a"
/complement(join(12895..12950,13560..13955,17307..17349))
/gene="Y47D3B.2a"
/notes="cDNA EST yk265b3.5 comes from this gene; cDNA EST
EMBL:C10452 comes from this gene; cDNA EST EMBL:C12093
comes from this gene; cDNA EST EMBL:D70235 comes from this
gene; cDNA EST yk552e6.3 comes from this gene"
/codon_start=1
/protein_id="CAA21040.1"
/db_xref="GI:3881006"
/translation="MRSLSFTLFFGLAALVWVNAQYTSLEDEDERGARGAMLHR
GGAFSADVGDDYKRGARAFYDEKRGARAFTEMKRGARVQFGEDEKRGARGA
FMKDRGGGARGAFDMKRGARAFVENSKRDEDWVIRPFEDDRLEKRGGRSFVYKPG
RLDD"
complement(13549..17349)
/gene="Y47D3B.2b"
/complement(join(13549..13955,17307..17349))
/gene="Y47D3B.2b"
/codon_start=1
/protein_id="CAB54383.1"
/db_xref="GI:5824792"
/translation="MRSLSFTLFFGLAALVWVNAQYTSLEDEDERGARGAMLHR
GGARFADVGDDYKRGARAFYDEKRGARAFTEMKRGARVQFGEDEKRGARGA
FMKDRGGGARGAFDMKRGARAFVENSKRDEDWVIRPFEDDRLEKRGGRSFVYKPG
RLDD"
complement(13549..17349)
/gene="Y47D3B.3"
/complement(join(20481..20514,21742..22022,22112..22190,23753..23852,
24487..24611,24655..24756,25563..25693,25796..25891,
27970..28341,29197..29332,29684..29742,29801..30076))
/gene="Y47D3B.3"
/notes="predicted using GeneFinder"
/codon_start=1
/protein_id="CAA21038.1"
/db_xref="GI:3881004"
/translation="MKNKILKVRSGLFIQGNFLISKYLRNTCISLDPQLIDLKLOT
RMVVPATFAIIVFFNFSLFNHTYISHYATLSLSFSFFFRHIFPHLPQOSTSD
QEVFLIYVLLVIVRSIAKDYDIEILPENCDCLDWYKACKVNGKWDVQYWKYC
DKERSDATFTGCGNGDVAVGONTTIEDLMVOCTODAKLHVEIEPFCEFNCTKR
VGKEYQGSFOWLCISLGRWITGCTYENETHDLLRVDHAYNGLHIEHVCRRQEPY
AIYVYQVRDVKLPRTNKGVRNRPPELQKIDEDINTFISLFCFVSCNTEGNS
TITPPIHSTSRAPLKHTRCHPSCDVRSGICHTILNLSRDXHCPACEVCRSK
LNFNEKYGKENDCAGGCMGRCVKNKDCSDCAVERAICKYIEQIFEFVYS
LNFNEKYGKENDCAGGCMGRCVKNKDCSDCAVERAICKYIEQIFEFVYS
DYEDNFAEKDEIFVKNYKLMONGRAPRVVEKKNHVEESYWCFLSPKNVKKRPS
WLKNAGLEVYSAYFDDRENSLFPNSAIQV"
30854..32478
/gene="Y47D3B.4"
/complement(join(30854..30889,31202..31345,31929..32127,32258..32478))
/gene="Y47D3B.4"
/notes="predicted using GeneFinder"
/codon_start=1
/protein_id="CAA21037.1"
/db_xref="GI:3881003"
/translation="MPMCLHAFOLLEDISGRLEWTEMOTLFGADSVSYHFFVLTNQ
THKVDYKHTNLKVIPLSLPSGNPNPSPIRSKYLKRNPKRRHLLPYNDCFYNY
ITHRYVLIIDIDEVIVPIQDNYGDLDRDFEAKSFYRLSSISRNRYKFPNSFIIF
```

```

gene
CDS
    join(34913..35016,38489..38669,39530..39640,39697..39824,
    39984..40076,40805..40983,41034..41172,42063..42383,
    42356..42498,42550..42774,43192..43259,43308..43485,
    43338..43666,44148..44300,44351..44419,44474..44658,
    44702..44801,45577..45675,45727..46265,47093..47246,
    47606..47707,48365..48475)
/gene="Y47D3B.5"
/notes="cDNA EST yk449e10.5 comes from this gene; cDNA EST
EMBL:D35319 comes from this gene; cDNA EST EMBL:C12322
comes from this gene; cDNA EST EMBL:D32633 comes from this
gene"
/codon_start=1
/protein_id="CAA21041.1"
/db_xref="GI:3881007"
/translation="MENPSTPLQASPTTLLADGYTDDLPALPSKYDNFRASRPYSRES
LUNTSLSGKVRFSQSFOSRDLTTIDDEEEYKPPAPETKFLQVSRRTVRLMA
GFSRAGSLEGVETGIVGPPMSPISFLGCKHYDYKLDRIASLVLYSFLGAW
VYLFEDHYEREVKLERIDRLMLNDFORISSVFRQOGLANFEVFDYKKLLHW
VRLPECLDWDYWGALFYVGLTFTTIGYNTYPTALGRAASVYVAVIGLPLAILSK
CGKMTDLSSEKQWQHRIQITKAKMTRKLRGKLLKSGEIVEANTGABADEKPEE
VESRTIPVIALILICVYVAGCSLFLWETWTFSTLFFCISLTLTIGLDIVPDK
PHMIFVYVIVGLSIVSFISVQIKIEECLRMIEQIKIEYMKKGLVDPEEIKR
SMACDPMNSHLKPAEQHTQWVRVDETEGPRNVEELKLORDDATSISRNDFRDI
SADPVLVSFAVKIMSGQNMKDTIKKDRLESKEITQTFPVKGMATVQEKYEQ
SDATSLPMSMSSPVAKKRNKTECAVCCDLPMTGAEFKADGTQDTLDAQFIDE
IAKLANLQTVRVPPIVEMATSAFMEDESLNSEALDPLRGLSQAINVICAMSSSF
FRNSQALKQIDIGVKDIETTSKQSLSLTPRLNLSIATSPVVSCLDEKLA
SKSKSRSSISSPPIPNLHEETETPLQNMVHMATDPTCSQVHRSTVTPVSGRE
DAETEMPQKMDRVSPVSTAIREDQNTSPVEQTPDFDRIQTSISEWFGKLLKKRD
ESQOTSICDDVVGSKKGGKASDENSQASSLGSVDFTSNPSSTSDSEKNNKK
IKKTLIAPNMTASVNSTOYSPPPGDFRLSEPHSTRASSTSGMSTSFEDFETREY
IVQTDLSYLKIARLDEYNNKTQFLPVCAAPLSSKEVEPFKSDRPSERSHYFGGG
IRASLGRKKPKQDMRIGESQTGQMDAELLKEVLSSEKRSISPRPCPNKRGVLJR
HPSLPTGIRSGKVSDYVIOHEKGIHNPATNRQSPKIRIQYSLDHDVI"
51392..52342
/gene="Y47D3B.6"
/join(51392..51449,51544..51641,51689..52342)
/gene="Y47D3B.6"
/notes="predicted using GeneFinder; cDNA EST yk506e5.3
comes from this gene"
/codon_start=1
/protein_id="CAA21043.2"
/db_xref="GI:5824793"
/translation="MERLICATLLAFTAAQVPVAPLYSACSAVPCAAPVCAQASVC
IFRAAPPMMAPPAEPAYAPAPLPAFPVFAAPVFAAPVFAAPVFAAPVFAAPV
IAAPALGCAAPCLGAACQACACACAPVSPVIGQPAIPVAPVAPVAGRLIPQ
ALPGSPCEPVECTGSGVGLCPPELIDEGTVCVARTIYGVPPVPPVPPVAPVIP
VALGAPCAAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV
complement(52583..66730)
/gene="Y47D3B.7"
/complement(join(52583..52774,52830..53190,61122..61275,
61322..61549,61609..62005,62315..62843,62889..63082,
63606..63841,64186..64605,64659..64767,65539..65848,
66107..66239,66652..66730))
/gene="Y47D3B.7"
/notes="predicted using GeneFinder; similar to
Helix-loop-helix DNA-binding domain; cDNA EST EMBL:D34815
comes from this gene; cDNA EST EMBL:D35004 comes from this
gene; cDNA EST EMBL:D68853 comes from this gene; cDNA EST
EMBL:D67853 comes from this gene; cDNA EST yk349h3.5 comes
```

```

alignment_scores:
    Quality: 176.00      Length: 176
    Ratio: 1.000        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x CEY47D3B/rev ..
```


Align seg 1/1 to reverse of: CEY47D3B from: 1 to: 95968

495 LeuLeuAlaGlySerAlaIlePheSerLysAlaAlaAlaGluAlaProI 511
|||||
62842 TTGCTCGCTGGAAGTCGGATATTCTCAAAAGCGCTGCAAGACTCCGAT 62793
511 eAlaSerProPheGluHisGlyArgValIleAspAspProAspGlyThrS 528
|||||
62792 TGGCTCCCGTTCGACATCGGAAGAGTGATGTGACCCGGATGGAACATA 62743
528 exThrArgThrLeuPheTrpGluGlySerIleIleAsnMetSerTyVal 544
|||||
62742 GCACCTCGGACGCTTTCTCGGAAGGAGTATCATCAATATGAGCTATGTC 62693
545 TrpValPheAsnIleLeuMetIleIleTyValValValValValVal 561
|||||
62692 TGGGTGTTCAACATCTTAATGATCATATATGTGTTGTCAAACTGCTGAT 62643
561 eHisGlyAspProValGlnAspPheMetSerValSerTrpGlnThrPheV 578
|||||
62642 CCATGGTGACCTTCTCAAGACTTCATGTCGGTTTCATGGCAGACTTTTG 62593
578 alThrThrArgGluLysAlaArgAlaGluLeuAsnSerGlyAsnLeuLys 594
|||||
62592 TGACGACTCGAGAGAGCGGAGCGAGTTGAACCTCTGGAATTTTGAAA 62543
595 AspAlaGlnArgLysPheCysGluCysLeuAlaThrLeuAspArgSerLe 611
|||||
62542 GATGCTCAGAGAAAGTCTCGAGAGTCTTGCACGCTTGCATGATCGCTG 62493
611 uProSerProGlyValAspSerValPheSerValGlyTrpGluCysValA 628
|||||
62492 TCCATCACCAGGGGGTGTGTTGCGGTGTTTGGGTGGGAAATGCGTTT 62443
628 rGHisLeuLeuAsnTrpLeuTrpIleGlyArgTyIleAlaArgArgArg 644
|||||
62442 GACATCTTTGATTGTTGTTGATCGGGAGATACATCGCAAGAGCGC 62393
645 ArgSerThrThrLysProValSerValValCysArgSerHisAlaGlnTh 661
|||||
62392 AGGTCCACCAGGAAGCTCTCTCAGTCGTTTGTAGGAGTCATCGCGACAG 62343
661 rAlaValLeuTyThrHisGluIleHisGln 670
|||||
62342 TGCAGTCTCTATCATCAATAATTTCATCAG 62315

seq_name: gb_htgl:CEY66A7

seq_documentation_block:
LOCUS CEY66A7 197735 bp DNA HTG 03-DEC-1998
DEFINITION Caenorhabditis elegans chromosome III clone Y66A7, *** SEQUENCING
IN PROGRESS ***, in unordered pieces.
ACCESSION AL022282
VERSION AL022282.1 GI:3451532
KEYWORDS HTG; HTGS-PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 197735)
AUTHORS Matthews,L.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwnematode.wustl.edu
COMMENT On Aug 25, 1998 this sequence version replaced gi:3250755.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate

segments:

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source 1. 197735 Location/Qualifiers
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="Y66A7"
BASE COUNT 62861 a 35824 c 35278 g 62783 t 989 others
ORIGIN

alignment_scores:
Quality: 176.00 Length: 176
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-332-522B-2 x CEY66A7/rev ..
Align seg 1/1 to reverse of: CEY66A7 from: 1 to: 197735

495 LeuLeuAlaGlySerAlaIlePheSerLysAlaAlaAlaGluAlaProI 511
|||||
3796 TTGCTCGCTGGAAGTCGGATATTCTCAAAAGCGCTGCAAGACTCCGAT 3747
511 eAlaSerProPheGluHisGlyArgValIleAspAspProAspGlyThrS 528
|||||
3746 TGGTCTCCCGTTCGACATCGGAAGAGTGATGATGACCCGGATGGAACATA 3697
528 exThrArgThrLeuPheTrpGluGlySerIleIleAsnMetSerTyVal 544
|||||
3696 GCACCTCGGACGCTTTCTGGAAGGAGGATCATCAATATGAGCTATGTC 3647
545 TrpValPheAsnIleLeuMetIleIleTyValValValValValVal 561
|||||
3646 TGGGTGTTCAACATCTTAATGATCATATATGTTGTGTTGTTGTTGTTGTT 3597
561 eHisGlyAspProValGlnAspPheMetSerValSerTrpGlnThrPheV 578
|||||
3596 CCATGGTGACCTTCTCAAGACTTCATGTCGGTTTCATGGCAGACTTTTG 3547
578 alThrThrArgGluLysAlaArgAlaGluLeuAsnSerGlyAsnLeuLys 594
|||||
3546 TGACGACTCGAGAGAGCGGAGCGAGTGTGAACCTCTGGAATTTTGAAA 3497
595 AspAlaGlnArgLysPheCysGluCysLeuAlaThrLeuAspArgSerLe 611
|||||
3496 GATGCTCAGAGAAAGTCTCGAGAGTCTTGCACGCTTGCATGATCGCTG 3447
611 uProSerProGlyValAspSerValPheSerValGlyTrpGluCysValA 628
|||||
3446 TCCATCACCAGGGGGTGTGTTGCGGTGTTTCCGGTTGGCTGGGAATGCGTTC 3397
628 rGHisLeuLeuAsnTrpLeuTrpIleGlyArgTyIleAlaArgArgArg 644
|||||
3396 GACATCTTTGATTGTTGTTGATCGGAGATACATCGCAAGAGCGCG 3347
645 ArgSerThrThrLysProValSerValValCysArgSerHisAlaGlnTh 661
|||||
3346 AGGTCCACCAGGAAGCTCTCTCAGTCGTTTGTAGGAGTCATCGCGACAG 3297
661 rAlaValLeuTyThrHisGluIleHisGln 670
|||||
3296 TGCAGTCTCTATCATCAATAATTTCATCAG 3269

seq_name: gb_htgl:CEY47D3

seq_documentation_block:

LOCUS CEY47D3 337565 bp DNA HTG 02-SEP-1999

```

DEFINITION Caenorhabditis elegans chromosome III clone Y47D3, *** SEQUENCING
IN PROGRESS ***, in unordered pieces.
ACCESSION Z98865
VERSION Z98865.1 GI:4164282
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 337565)
AUTHORS Matthews,L.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1999) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwenematode.wustl.edu
COMMENT
On Jan 19, 1999 this sequence version replaced gi:4056520.
Order of segments is not known; 800 n's separate segments.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES             Location/Qualifiers
     source            1..337565
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="III"
                     /clone="Y47D3"
BASE COUNT 108087 a 59856 c 60930 g 107892 t      800 others
ORIGIN

alignment_scores:
  Quality: 176.00      Length: 176
  Ratio: 1.000         Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522b-2 x CEY47D3/rev ..
Align seg 1/1 to reverse of: CEY47D3 from: 1 to: 337565
495 LeuLeuAlaGlySerAlaIlePheSerLysAlaAlaAlaGluAlaProil 511
|||||
304439 TTGCTCGCTGGAGTGGCATATCTCAAAAGCGCTGCAGAACGTCGGAT 304390
|||||
511 eAlaSerProPheGluHisGlyArgValIleAspAspProAspGlyThrs 528
|||||
304389 TGCCTCCCGTTCGAGCATGGAAGAGTGATTGATGACCGGATGGAAC 304340
|||||
528 erThrArgThrLeuPheTrpGluGlySerIleIleAsnMetSerTyVal 544
|||||
304339 GCATCGGACGCTTTCTGGGAAGGGAGTATCATCAATAGCATATGTC 304290
|||||
545 TrpValPheAsnIleLeuMetIleIleTyValValValLysLeuLeuIle 561
|||||
304289 TGGGTGTTCAACATCTTAATGATCATATATGTTGGTGTCAAACTGCTGAT 304240
|||||
561 eHisGlyAspProValGlnAspPheMetSerValSerTrpGlnThrPheV 578
|||||
304239 CCATGGTGACCCGTTCAGACATTCATGTCCTGTTTCATGGCAGACTTTG 304190
|||||
578 alThrThrArgGluLysAlaArgAlaGluLeuAsnSerGlyAsnLeuLys 594
|||||
304189 TGACGACTCGAGAGAGCGGAGAGCCGAGTTGAACCTGGAAATTTGAA 304140
|||||
595 AspAlaGlnArgLysPheCysGluCysLeuAlaThrLeuAspArgSerLe 611
|||||

```

```

|||||
304139 GATGCTCAGAGAAAGTCTGCGAGTGTCTGCAACGTTGGATCGATCGCT 304090
|||||
611 uProSerProGlyValAspSerValPheSerValGlyTrpGluCysVala 628
|||||
304089 TCCATCACCGGGGTTGATTGCGTGTTCGTTGCGTGGGAATGCGTTC 304040
|||||
628 rGHisLeuLeuAsnTrpLeuTrpIleGlyArgTyTrileAlaArgArgArg 644
|||||
304039 GACATCTTTTGAATTGTTGTGATCGGGAGATACATCGCAAGAGCGC 303990
|||||
645 ArgSerThrThyLysProValSerValValCysArgSerHisAlaGlnTh 661
|||||
303989 AGGTCCACCACAGAGCTGTCAGTCGTTGTAGAGTCATGCGCAGAC 303940
|||||
661 rAlaValLeuTyHisGluIleHisGln 670
|||||
303939 TGCAGTCTCTATCATGAAATTCATCAG 303912
|||||
seq_name: gb_htgl1.CEH10N23
seq_documentation_block:
LOCUS CEH10N23 36780 bp DNA HTG 22-SEP-1998
DEFINITION Caenorhabditis elegans chromosome III clone H10N23, *** SEQUENCING
IN PROGRESS ***, in unordered pieces.
ACCESSION Z94157
VERSION Z94157.1 GI:3377969
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 36780)
AUTHORS Sulston,J.
TITLE Direct Submission
JOURNAL Submitted (21-SEP-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwenematode.wustl.edu
COMMENT
On Aug 3, 1998 this sequence version replaced gi:1945146.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES             Location/Qualifiers
     source            1..36780
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="III"
                     /clone="H10N23"
BASE COUNT 8697 a 5380 c 5339 g 8564 t 8800 others
ORIGIN

alignment_scores:
  Quality: 153.00      Length: 153
  Ratio: 1.000         Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522b-2 x CEH10N23/rev ..
Align seg 1/1 to reverse of: CEH10N23 from: 1 to: 36780
495 LeuLeuAlaGlySerAlaIlePheSerLysAlaAlaAlaGluAlaProil 511
|||||

```

```

461 TTGCTCGTGGAGTGCATATATCTCAAAAGCCGCTGCAGAAGCTCCGAT 412
511 eaIaSerProPheGluHisGlyArgValIleAspSerProAspGlyThrS 528
411 TGCCCTCCCGCTTCGAGCATGGAGAGTATTGATACCCGGATGGAACATA 362
528 exThrArgThrLeuPheTrpGluGlySerIleIleAsnMetSerTyrrVal 544
361 GCACCTCGAGCGCTTCTGGGAAGGAGATCATCAATATGAGCTATGTC 312
545 TrpValPheAsnIleLeuMetIleIleTyrrValValValLysLeuLeuI 561
311 TGGGTGTTCAACATCTTAATGATCATATATGTTGTCAAACTGCTGAT 262
561 eHisGlyAspProValGlnAspPheMetSerValSerTrpGlnThrPheV 578
261 CCATGGTGACCTGTTCAGAGCTTCATGTCCTTCATGGCGACGCTTTTG 212
578 alThrThrArgGluLysAlaArgAlaGluLeuAsnSerGlyAsnLeuLys 594
211 TCACGACTCGAGAGAGCGGAGAGCGGAGTTGAATTCGGAATTTGAAA 162
595 AspAlaGlnArgLysPheCysGluCysLeuAlaThrLeuAspArgSerLe 611
161 GATGCTCAGAGAAAGTTCTCGAGTGTCTTGCAACGTTGGATCGATCGCT 112
611 uProSerProGlyValAspSerValPheSerValGlyTrpGluCysValA 628
111 TCCATCACCGGGGTGATTTCGGTGTCTTCGGTGGCTGGGAATCGCTTC 62
628 tGHisLeuLeuAsnTrpLeuTrpIleGlyArgTyrrIleAlaArgArg 644
61 GACATCTTTGAAATGGTTGGTGGATCGGAGATACATCGCAAGAGGCC 12
645 ArgSerThr 647
11 AGGTCCACC 3

```

seq_name: gb_htg1:CEH10N23

seq_documentation_block:
LOCUS CEH10N23 36780 bp DNA HTG 22-SEP-1998
DEFINITION Caenorhabditis elegans chromosome III clone H10N23, *** SEQUENCING
IN PROGRESS ***, in unordered pieces.

ACCESSION 294157
VERSION 294157.1 GI:3377969
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans

REFERENCE 1 (bases 1 to 36780)
AUTHORS Sulston, J.
JOURNAL Direct Submission

COMMENT Submitted (21-SEP-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwenematode.wustl.edu
On Aug 3, 1998 this sequence version replaced gi:1945146.

IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source Location/Qualifiers
1..36780 /organism="Caenorhabditis elegans"

```

/db_xref="taxon:6239"
/chromosome="III"
/clone="H10N23"
BASE COUNT 8697 a 5380 c 5339 g 8564 t 8800 others
ORIGIN

```

alignment_scores:
Quality: 126.00 Length: 126
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x CEH10N23

Align seg 1/1 to: CEH10N23 from: 1 to: 36780

```

211 AlaGlnGlyProSerGlyProSerTyrrSerGlnHisHisGlnSerPr 227
|||||
6088 GCTCAAGGCCCATCAGGACCATCATATTACTACACACCATCATCTCC 6137
|||||
227 OProProHisHisHisHisHisHisHisHisHisHisHisHisHis 244
|||||
6138 ACCACCTCATCACCACCATCACCACCCGATGCCAAAATCCATGAGA 6187
|||||
244 roGluGlnValAlaSerProSerIleGluAspAlaProGluThrLysPro 260
|||||
6188 CTGAACAAGTGGCATCTCCATCGATTGAAGATGTCTCCAGACGACGA 6237
|||||
261 ThrHisLeuValGluProGlnSerProLysSerProGlnAsnMetLysG 277
|||||
6238 ACTCATTTGGTTGAACCAACAAAGTCCAAAAGCCCGCAGATATGA 6287
|||||
277 uGluLeuLeuArgLeuLeuValAsnMetSerProSerGluValGluArg 294
|||||
6288 GGAGCTTCTTCGGTTACTAGTTAACATGCTCTCCGAGTGAAGTTGA 6337
|||||
294 euLysAsnLysLysSerGlyAlaCysSerAlaThrAsnGlyProSerArg 310
|||||
6338 TAAAGAATAAAAAATCAGGAGCATGTTTCAGCGACGAATGGGCCATC 6387
|||||
311 SerLysGluLysAlaAlaLysIleValIleGlnGluThrAlaGluGly 327
|||||
6388 AGTAAGGAGAGAGCGCGGCAAGATTGTGATTGAGGAGACACCGCA 6437
|||||
327 pGluAspGluAspAspGluAspSerAsp 336
|||||
6438 TGAAGATCAGGATGATGAGGATAGTGAT 6465
|||||

```

seq_name: gb_htg1:CEY48A6

seq_documentation_block:

LOCUS CEY48A6 296699 bp DNA HTG 31-JUL-1998
DEFINITION Caenorhabditis elegans chromosome III clone Y48A6, *** SEQUENCING
IN PROGRESS ***, in unordered pieces.

ACCESSION 292854
VERSION 292854.1 GI:3218061
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans

REFERENCE 1 (bases 1 to 296699)
AUTHORS Gardner, A.
TITLE Direct Submission

JOURNAL Submitted (30-JUL-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwenematode.wustl.edu
On Jun 13, 1998 this sequence version replaced gi:3204147.

Order of segments is not known; 800 n's separate segments.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress

and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.

* NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 source Location/Qualifiers
 1..296699
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="III"
 /clone="Y48A6"
 BASE COUNT 93035 a 52118 c 54334 g 90807 t 6405 others
 ORIGIN

alignment_scores:
 Quality: 120.00 Length: 120
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x CEY48A6 ..
 Align seg 1/1 to: CEY48A6 from: 1 to: 296699

931 AsnLeuGlyLeuAlaValGlyHisAlaLeuCysAlaArgLysIleCys11 947
 |||||
 19489 AATCTGGTTTGGCGTGGCCACGGTGTGCTCGCAAGATTGCAAT 19538

947 eAspAspArgAspSerProLysValSerGlnTyrValCysIleHisThrL 964
 |||||
 19539 AGATGACCGAGATTCCGAAAGTCAGTCAATACGTGTGCATTCACAAA 19588

964 yLysSerLeuGluSerLeuArgLeuPheSerThrSerArgAlaSer 980
 |||||
 19589 AGAAGTCGCTCGAATCCCTCCGACTATTCCACATCATCGCGAGCATCA 19638

981 GlyValValSerGlyIleGlnGluGlyThrArgArgMetAlaTyrGluTr 997
 |||||
 19639 GGTGTGCTGCTCGAATTCAGGAAGGTACACGCCGAATGCCCTACGAATG 19688

997 pIleMetAsnSerLeuLeuAspAlaTrpArgSerAsnLeuPheAlaSerL 1014
 |||||
 19689 GATTATGAATCGCTGCTGACGCGTGGCGTTCCAAATCTATTGCGATCGA 19738

1014 ySProTyrTrpThrGlnSerPheLysGlyGlnSerThrPheSerThrLeu 1030
 |||||
 19739 AACCCCTACTGGACACAAAGCTTCAAGGACAAATCCACGTTTATGATCGCTT 19788

1031 TyrGlnGluAlaTyrAsnHisTyrAlaIleIleAsnGlyThrArgGlyAs 1047
 |||||
 19789 TATCAAGAGCGGTATATCATTTATGCGATTATTAATGGACACAGGGGAGA 19838

1047 pCysTrpArg 1050
 |||||
 19839 TTGTTGGAGA 19848

seq_name: gb_pr3:AF039153
 seq_documentation_block:
 LOCUS AF039153 732 bp DNA PRI 02-JUN-1998
 DEFINITION Homo sapiens chromosome 4 cosmid B2 subclone 1, D424 repeat
 sequence.
 ACCESSION AF039153
 VERSION AF039153.1 GI:3170147
 KEYWORDS

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 732)
 AUTHORS Dickson,M.C., Heather,L.J., Bolland,D.J., van Geel,M., de Jong,P., Flint,J., Frants,R.R. and Hewitt,J.E.
 TITLE Inter- and intrachromosomal duplications of human chromosome 4q35: an example of telomere plasticity
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 732)
 AUTHORS Dickson,M.C.
 TITLE Direct Submission
 JOURNAL Submitted (18-DEC-1997) School of Biological Sciences, 3.239 Stopford Building, University of Manchester, Oxford Road, Manchester,M13 9PT, United Kingdom

FEATURES
 source Location/Qualifiers
 1..732
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4q35"
 /clone="cosmid B2 derived from YAC RM2173"
 /sub_clone="1"
 /note="BamHI fragment"
 misc_feature 1..408
 repeat_region <1..>732
 /rpt_type=tandem
 /rpt_family="D424"
 misc_feature 475..732
 /note="partial double homeobox"

BASE COUNT 107 a 306 c 224 g 95 t
 ORIGIN

alignment_scores:
 Quality: 10.00 Length: 10
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x AF039153 ..
 Align seg 1/1 to: AF039153 from: 1 to: 732

227 ProProProHisHisHisHisHisHisPro 236
 |||||
 129 CCCCCCCCCCACCACCCACCACCCACCC 158

seq_name: gb_ph:PT4G67G21
 seq_documentation_block:
 LOCUS PT4G67G21 767 bp DNA PHG 28-APR-1993
 DEFINITION Bacteriophage T4 genes 67, 68 and 21.
 ACCESSION J02512 J02505
 VERSION J02512.1 GI:215909
 KEYWORDS head core protein
 SOURCE Bacteriophage T4 DNA.
 ORGANISM coliphage T4
 Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
 T4-like phages.

REFERENCE 1 (bases 1 to 337)
 AUTHORS Voelker,T.A., Gafner,J., Bickler,T.A. and Showe,M.K.
 TITLE Gene 67, a new, essential bacteriophage T4 head gene codes for a
 prehead core component, pIP: I. Genetic mapping and DNA sequence
 J. Mol. Biol. 161, 479-489 (1982)

REFERENCE 2 (bases 379 to 434)
 AUTHORS Christensen,A.C. and Young,E.T.
 TITLE T4 late transcripts are initiated near a conserved DNA sequence
 JOURNAL Nature 299, 369-371 (1982)
 MEDLINE 82272427

REFERENCE 3 (bases 246 to 767)
 AUTHORS Keller,B., Sengstag,C., Kellenberger,E. and Bickle,T.A.
 TITLE Gene 68, a new bacteriophage T4 gene which codes for the 17K
 prehead core protein is involved in head size determination

J. Mol. Biol. 179, 415-430 (1984)
 MEDLINE 85083058 Location/Qualifiers
 FEATURES
 source 1. .767
 /organism="coliphage T4"
 /db_xref="taxon:10665"
 CDS 57..299
 /note="prehead core protein PIP (g67)"
 /codon_start=1
 /transl_table=11
 /protein_id="AAA32520.1"
 /db_xref="GI:215910"
 /translation="MEGLTEAKNSDLVAARKLFAFAMARTIDLIKEEKIAIARNEL
 IGEPEDEDEDEDESDKDDKDEDEDEDE"
 mat_peptide 135..296
 CDS 299..724
 /note="internal peptide II"
 /note="17k prohead core protein (g68)"
 /codon_start=1
 /transl_table=11
 /protein_id="AAA32521.1"
 /db_xref="GI:215911"
 /translation="MLLIPETHELVLNVEALIPAEQGRFDELSSALNKDDINTIVEN
 MLDDETDLAVASINENMPLNEFIVKHSARGEITRTDKRTREARNAFQTTGLSKAK
 RQIARKATKTIANPAGQSRQKRKKKALKRKGLS"
 mat_peptide 362..721
 CDS 724..>767
 /note="g21 protein"
 /codon_start=1
 /transl_table=11
 /protein_id="AAA32522.1"
 /db_xref="GI:215912"
 /translation="MNEPOLLITETGKPG"
 BASE COUNT 275 a 140 c 163 g 189 t
 ORIGIN 75 bp upstream of AluI site; 101.6 kb on genomic map.

alignment_scores
 Quality: 10.00 Length: 10
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x PT4G67G21 ..

Align seg 1/1 to: PT4G67G21 from: 1 to: 767

327 AspGluAspGluAspAspGluAspSerAsp 336
 |||||
 216 GATGAAGATGAGATGACGAGATAGTAGT 245

seq_name: gb_p11:AB000209

seq_documentation_block:
 LOCUS AB000209 1059 bp DNA PLN 08-APR-2000
 DEFINITION Nannochloropsis oculata mitochondrial DNA for cytochrome oxidase subunit I, partial cds.
 ACCESSION AB000209
 VERSION AB000209.1 GI:2913874
 KEYWORDS cytochrome oxidase subunit I.
 SOURCE Nannochloropsis oculata (strain:CCAP 849/1) mitochondrion DNA.
 ORGANISM Eukaryota; stramenopiles; Eustigmatophyceae; Nannochloropsis.
 REFERENCE 1 (sites)
 AUTHORS Ehara,M., Hayashi-Ishimaru,Y., Inagaki,Y. and Ohama,T.
 TITLE Use of a deviant mitochondrial genetic code in yellow-green algae as a landmark for segregating members within the phylum
 J. Mol. Biol. 179, 415-430 (1984)
 JOURNAL 97383241
 MEDLINE 97383241
 REFERENCE 2 (sites)
 AUTHORS Inagaki,Y., Hayashi-Ishimaru,Y., Ehara,M., Igarashi,I. and Ohama,T.
 TITLE Algae or protozoa: phylogenetic position of euglenophytes and

dinoflagellates as inferred from mitochondrial sequences
 J. Mol. Evol. 45 (3), 295-300 (1997)
 MEDLINE 97451042
 REFERENCE 3 (sites)
 AUTHORS Hayashi-Ishimaru,Y., Ehara,M., Inagaki,Y. and Ohama,T.
 TITLE A deviant mitochondrial genetic code in prymnesiophytes (yellow-algae): UGA codon for tryptophan
 Curr. Genet. 32 (4), 296-299 (1997)
 JOURNAL 98004465
 MEDLINE 98004465
 REFERENCE 4 (bases 1 to 1059)
 AUTHORS Ehara,M.
 TITLE Direct Submission
 JOURNAL Submitted (29-DEC-1996) to the DDBJ/EMBL/GenBank databases. Megumi Ehara, Osaka Univ., Department of Biology; Machikaneyama-cho 1-1, Toyonaka, Osaka 560, Japan (E-mail:ehara@ims.brh.co.jp, Tel:+81-726-81-9761, Fax:+81-726-81-9757)

FEATURES
 Location/Qualifiers
 1..1059
 /organism="Nannochloropsis oculata"
 /strain="CCAP 849/1"
 /db_xref="taxon:43925"
 gene 1..1059
 /gene="COXI"
 CDS <1..>1059
 /gene="COXI"
 /codon_start=1
 /product="cytochrome oxidase subunit I"
 /protein_id="BAA24973.1"
 /db_xref="GI:2913875"
 /translation="GTLLIFGAIAGVAGTTLSILIRLEAQPGNQFLSGNNQLNVI
 VIGHAFVMIFFVPLVIGGFGNFWPLMIGAPDMAPRMNNISFWLPPSLILLAS
 TEVEAGAGTGTVPPLSGAHSQSPVDLAIFSLHSGAASILGAINFITTFNNRA
 PGMNHRPLFLFVWSYLITAFLLLSLPVFAAGITMLLTDRNFTTFYDPAGGDPVLY
 QHLFWFHPHYIILIPAFGIISHVSSFSNKPVEGYLGMIVAILSIGVGFIVWAH
 HMTVGLDIDTRAYTAATMIATVPTGKIFSWATMWSGFELKTPMLFAIGVISLF
 TVGGYGVVLANGIDVAL"
 BASE COUNT 228 a 187 c 214 g 430 t
 ORIGIN

alignment_scores
 Quality: 10.00 Length: 10
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x AB000209 ..

Align seg 1/1 to: AB000209 from: 1 to: 1059

818 LeuSerLysLeuValGlnGluLeuValGly 827
 |||||
 311 TTGTGGAAGCTGGTGCAGGAACTGGTTGA 340

seq_name: gb_bt5:AC014381

seq_documentation_block:
 LOCUS AC014381 12029 bp DNA HTG 16-NOV-1999
 DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
 ACCESSION AC014381
 VERSION AC014381.1 GI:6436954
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 12029)
 AUTHORS Adams,M. and Venter,J.C.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT This sequence was identified as CDM:10209537 by the submitter.
 For further information on this sequence e-mail to fly@celera.com.

- * NOTE: This is a 'working draft' sequence.
- * This sequence will be replaced
- * by the finished sequence as soon as it is available and
- * the accession number will be preserved.

FEATURES
 source
 1..12029
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 BASE COUNT 3251 a 2649 c 2685 g 3444 t
 ORIGIN

alignment_scores:
 Quality: 10.00 Length: 10
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x AC014381/rev ..
 Align seg 1/1 to reverse of: AC014381 from: 1 to: 12029

227 ProProHisHisHisHisHisPro 236
 9277 CCCCCCCCCATCACACCACCCACCC 9248

seq_name: gb_pr4:AF117653

seq_documentation_block:
 LOCUS AF117653 13479 bp DNA PRI 10-SEP-1999
 DEFINITION Homo sapiens double homeobox protein genes, complete cds.
 ACCESSION AF117653
 VERSION AF117653.1 GI:5852424
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 13479)
 AUTHORS Gabriels,J., Beckers,M.C., Ding,H., De Vriese,A., Plaisance,S., van
 der Maarel,S.M., Padberg,G.W., Frants,R.R., Hewitt,J.E., Collen,D.
 and Belayew,A.
 TITLE Nucleotide sequence of the partially deleted D424 locus in a
 patient with FSHD identifies a putative gene within each 3.3 kb
 element

JOURNAL Gene 236 (1), 25-32 (1999)
 MEDLINE 99365298
 REFERENCE 2 (bases 1 to 13479)
 AUTHORS Gabriels,J.
 TITLE Direct Submission
 JOURNAL Submitted (04-JAN-1999) Center for Molecular and Vascular Biology,
 University of Leuven, Herestraat 49, Leuven B-3000, Belgium

FEATURES
 source
 1..13479
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="4q35; FSHD locus"
 5738..9053
 /rpt_family="D424"
 /rpt_type-dispersed
 repeat_region
 7311..7319
 /note="TACAA"
 GC_signal
 TATA_signal
 7385..7389
 /note="TACAA"
 CDS
 7534..8709
 /note="DUX4"
 /codon_start=1
 /product="double homeobox protein"
 /protein_id="AAD54067.1"
 /db_xref="GI:5852425"
 /translation="MALPTPSDSTLPAEARGRRRLVWTPSQSEALRACFERNPYP"

misc_feature
 /note="Region: homeobox domain"
 7813..7992
 /note="Region: homeobox domain"
 9035..12331
 /rpt_family="D424"
 /rpt_type-dispersed
 GC_signal
 TATA_signal
 10607..10615
 10681..10685
 /note="TACAA"
 CDS
 10830..12005
 /note="DUX4"
 /codon_start=1
 /product="double homeobox protein"
 /protein_id="AAD54068.1"
 /db_xref="GI:5852426"
 /translation="MALPTPSDSTLPAEARGRRRLVWTPSQSEALRACFERNPYP"
 GIATRLAQAGIPEPRVQINQFQNSRQLQHRRESRQHPGRRGPPGRRKRTAVT
 GSQTALLRAFEDKDRPFGIAAREELARETGLPAPFCAPGALPQGAFAVQAARAAPAL
 GGLCSAAGGGHPAPSWAFATGAWTGLPAPFCAPGALPQGAFAVQAARAAPAL
 OPSQAAPAGVSQAPARGDFAYAAPAPPEGRSPTLLGLLRTAKAGTGTGTSAT
 ACAPARWHSILGPKRGRPRGACATHVPGESVGLPGSGPGRGGVGPSPRGSTSP
 ARAPGTPTPPRGRGRCKASRRPRRSRRRLGHPSPAACCCWSSWRARFCFCSRRLS"

misc_feature
 /note="Region: homeobox domain"
 1109..11288
 /note="Region: homeobox domain"
 BASE COUNT 2533 a 4317 c 3961 g 2668 t
 ORIGIN

alignment_scores:
 Quality: 10.00 Length: 10
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x AF117653 ..

Align seg 1/1 to: AF117653 from: 1 to: 13479

227 ProProHisHisHisHisHisPro 236
 7241 CCCCCCCCCACCCACCCACCC 7270

seq_name: gb_ph:AF158101

seq_documentation_block:
 LOCUS AF158101 168900 bp DNA PHG 16-JUL-1999
 DEFINITION Coliphage T4, complete genome.
 ACCESSION AF158101
 VERSION AF158101.3 GI:5508842
 KEYWORDS
 SOURCE coliphage T4.
 ORGANISM coliphage T4.

Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
 T4-like phages.
 REFERENCE 1 (bases 167962 to 168900)
 AUTHORS Priboj,D., Sigurdson,D.C., Gold,L., Singer,B.S., Napoli,C.,
 Brosius,J., Dull,T.J. and Noller,H.F.
 TITLE rII cistrons of bacteriophage T4. DNA sequence around the
 intercistronic divide and positions of genetic landmarks
 J. Mol. Biol. 149 (3), 337-376 (1981)
 JOURNAL 82078066
 MEDLINE
 REFERENCE 2 (bases 155865 to 159619)
 AUTHORS Oliver,D.B. and Crowther,R.A.
 TITLE DNA sequence of the tail fibre genes 36 and 37 of bacteriophage T4

regions
Gene 57 (1), 143-148 (1987)
88112857
Erratum:[published erratum appears in Gene 1988 May 30;65(2):337]]
27 (bases 8225 to 9150)
REFERENCE
Gaus, P., Gayle, M., Winter, R.B. and Gold, L.
The bacteriophage T4 dexa gene: sequence and analysis of a gene
conditionally required for DNA replication
Mol. Gen. Genet. 206 (1), 24-34 (1987)
JOURNAL
MEDLINE
87201086
28 (bases 33055 to 33290)
REFERENCE
Hsu, T., Wei, R.X., Dawson, M. and Karam, J.D.
Identification of two new bacteriophage T4
roles in transcription and DNA replication
J. Virol. 61 (2), 366-374 (1987)
JOURNAL
MEDLINE
87112930
29 (bases 40190 to 48400)
REFERENCE
Tomaschewski, J. and Ruger, W.
Nucleotide sequence and primary structures of gene products coded
for by the T4 genome between map positions 48.266 kb and 39.166 kb
Nucleic Acids Res. 15 (8), 3632-3633 (1987)
JOURNAL
MEDLINE
87203398
30 (bases 159646 to 160197)
REFERENCE
Montag, D., Riede, I., Eschbach, M.L., Degen, M. and Henning, U.
Receptor-recognizing proteins of T-even type bacteriophages.
Constant and hypervariable regions and an unusual case of evolution
J. Mol. Biol. 196 (1), 165-174 (1987)
JOURNAL
MEDLINE
88011316
31 (bases 160218 to 160874)
REFERENCE
Montag, D., Degen, M. and Henning, U.
Nucleotide sequence of gene t (lysis gene) of the E. coli phage T4
Nucleic Acids Res. 15 (16), 6736 (1987)
JOURNAL
MEDLINE
87316934
32 (bases 2458 to 3367)
REFERENCE
Huang, W.M., Ao, S.Z., Casjens, S., Orlandi, R., Zeikus, R., Weiss, R.,
alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-332-522b-2 x AF158101 ..
Align seg 1/1 to: AF158101 from: 1 to: 168900
327 AspGluAspGluAspAspGluAspSerAsp 336
|||||
102937 GATGAATGAGATGACGAGATAGTAT 102966
seq_name: gb_hgt21:AC011080
seq_documentation_block:
LOCUS AC011080 179921 bp DNA HTG 30-MAR-2000
DEFINITION Homo sapiens clone RP11-45K2, WORKING DRAFT SEQUENCE, 27 unordered
pieces.
ACCESSION AC011080
VERSION AC011080.2 GI:7341694
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179921)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens, clone RP11-45K2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179921)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Balwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhvalter, B.,
Brown, A., Castle, A., Collangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehocky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Sever, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 30, 2000 this sequence version replaced gi:6006259.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1174
Center clone name: 45_K2
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 151041 bases at least Q40
Consensus quality: 163342 bases at least Q30
Consensus quality: 169295 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 177321; sum-of-contigs
Quality coverage: 2.9 in Q20 bases; agarose-fp
Quality coverage: 3.1 in Q20 bases; sum-of-contigs

* NOTE: this is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1186: contig of 1186 bp in length
* 1187 1286: gap of 100 bp
* 1287 2524: contig of 1238 bp in length
* 2525 2624: gap of 100 bp
* 2625 3850: contig of 1226 bp in length
* 3851 3950: gap of 100 bp
* 3951 4972: contig of 1022 bp in length
* 4973 5072: gap of 100 bp
* 5073 6616: contig of 1544 bp in length
* 6617 6716: gap of 100 bp
* 6717 8945: contig of 2229 bp in length
* 8946 9045: gap of 100 bp
* 9046 11297: contig of 2252 bp in length
* 11298 11397: gap of 100 bp
* 11398 13918: contig of 2519 bp in length
* 13917 14016: gap of 100 bp
* 14017 19211: contig of 5195 bp in length
* 19212 19311: gap of 100 bp
* 19312 23451: contig of 4140 bp in length
* 23452 23551: gap of 100 bp
* 23552 27030: contig of 3479 bp in length
* 27031 27130: gap of 100 bp
* 27131 30176: contig of 3046 bp in length
* 30177 30276: gap of 100 bp
* 30277 34726: contig of 4450 bp in length
* 34727 34826: gap of 100 bp
* 34827 41337: contig of 6511 bp in length
* 41338 41437: gap of 100 bp
* 41438 47549: contig of 6112 bp in length


```

* 47550 47649: gap of 100 bp
* 47650 52582: contig of 4933 bp in length
* 52583 52682: gap of 100 bp
* 52683 58617: contig of 5935 bp in length
* 58618 58717: gap of 100 bp
* 58718 67895: contig of 9178 bp in length
* 67896 67995: gap of 100 bp
* 67996 76576: contig of 8581 bp in length
* 76577 76678: gap of 100 bp
* 76677 85661: contig of 8985 bp in length
* 85662 85761: gap of 100 bp
* 85762 95394: contig of 9633 bp in length
* 95395 95494: gap of 100 bp
* 95495 108165: contig of 12671 bp in length
* 108166 108265: gap of 100 bp
* 108266 120754: contig of 12489 bp in length
* 120755 120854: gap of 100 bp
* 120855 130257: contig of 9403 bp in length
* 130258 130357: gap of 100 bp
* 130358 140625: contig of 10268 bp in length
* 140626 140725: gap of 100 bp
* 140726 154151: contig of 13426 bp in length
* 154152 154251: gap of 100 bp
* 154252 179921: contig of 25670 bp in length.
FEATURES
    Location/Qualifiers
        1..179921
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="RP11-45K2"
            /clone_lib="RPC1-11 Human Male BAC"
        1..1186
            /note="assembly_fragment"
        1287..2524
            /note="assembly_fragment"
        2625..3850
            /note="assembly_fragment"
        3951..4972
            /note="assembly_fragment"
        5073..6616
            /note="assembly_fragment"
        6717..8945
            /note="assembly_fragment"
        9046..11297
            /note="assembly_fragment"
        11398..113916
            /note="assembly_fragment"
        14017..19211
            /note="assembly_fragment"
        19312..23451
            /note="assembly_fragment"
        23552..27030
            /note="assembly_fragment"
        27131..30176
            /note="assembly_fragment"
        clone_end:T7
        vector_side:left
        30277..34726
            /note="assembly_fragment"
        34827..41337
            /note="assembly_fragment"
        41438..47549
            /note="assembly_fragment"
        47650..52582
            /note="assembly_fragment"
        clone_end:SP6
        vector_side:right
        52683..58617
            /note="assembly_fragment"
        58718..67895
            /note="assembly_fragment"
        67996..76576
            /note="assembly_fragment"
        76677..85661
            /note="assembly_fragment"

```

```

/note="assembly_fragment"
85762..95394
/note="assembly_fragment"
95495..108165
/note="assembly_fragment"
108266..120754
/note="assembly_fragment"
120855..130257
/note="assembly_fragment"
130358..140625
/note="assembly_fragment"
140726..154151
/note="assembly_fragment"
154252..179921
/note="assembly_fragment"
BASE COUNT 55473 a 33850 c 33328 g 54664 t 2606 others
ORIGIN

alignment_scores:
    Quality: 10.00      Length: 10
    Ratio: 1.000        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-332-522b-2 x AC011080/rev ..

Align seg 1/1 to reverse of: AC011080 from: 1 to: 179921

964 LysLysSerLeuGluSerLeuArgLeuPhe 973
|||||
29404 AAAAGAGTTTGGAGTCACTTAGGCTATTT 29375

seq_name: gb_in2:AE003429

seq_documentation_block:
LOCUS AE003429 297405 bp DNA INV 22-MAR-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386054 section 13
of 35, complete sequence.
ACCESSION AE003429 AE002566
VERSION AE003429.1 GI:7290438
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 297405)
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Ananidis,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champs,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Mikos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Besson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkov,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
Fleischmann,W., Fosler,C., Gabriellian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Ma,J., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,

```

Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusskern,D.R., Pacleb,J.M., Palazolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svizskas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A., Weinstein,G.M., Weissbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.

The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)

20196006

2 (bases 1 to 297405)

ADAMS,M.D., CELNIKER,S.E., GIBBS,R.A., RUBIN,G.M. and VENTER,C.J.

Direct Submission

Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

FEATURES

source

1..297405

Location/Qualifiers

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/chromosome="X"

Join(2369..2585,2666..2787,3005..3651,3726..4013)

/gene="CG2947"

/product="CT9995"

/db_xref="FLYBASE:FBan0002947"

/db_xref="FLYBASE:FBgn0029676"

<2369..>4013

/gene="CG2947"

/db_xref="FLYBASE:FBan0002947"

/db_xref="FLYBASE:FBgn0029676"

Join(2425..2585,2666..2787,3005..3651,3726..3917)

/gene="CG2947"

/note="CG2947 gene product"

/codon_start=1

/db_xref="FLYBASE:FBan0002947"

/db_xref="FLYBASE:FBgn0029676"

/protein_id="AAF45894.1"

/db_xref="GI:7290439"

/translation="MQGDLKKLYFIDFALENFTFLNMPOLQVQVDFVEKFGTVPP
GQFNAGSGKCPFGVAGAKANEPANAPEDSEKSLSPESVDLEMDMEGVTEADSD
PAQPMGYSKATTEEVQSELRAQASAYGQKQFDEATLYTKAIELSPGNALFHA
KRGQAFLLKPNACIRDCVALLNSDLAAGYKFRGRARLLGDFLAADHLROACK
LDPEETDEWKEVTPNAKITEOHLKQERQAERKIKERODORRARKQEKHNASS
GSSGEFGGNGNGNSDILGMSDPEVSAIQDILSNPGNITKYASNPKIYNLIKK
IYPGDVGAATGQAGEKAGKPESEPKAKDSADFVDDGLD"

<5725..>6921

/gene="CG15238"

/product="CT35178"

/db_xref="FLYBASE:FBan0015238"

/db_xref="FLYBASE:FBgn0029677"

/evidence=not_experimental

<5725..>6921

/gene="CG15238"

/db_xref="FLYBASE:FBan0015238"

/db_xref="FLYBASE:FBgn0029677"

/evidence=not_experimental

5725..6921

/gene="CG15238"

/note="CG15238 gene product"

/codon_start=1

/db_xref="FLYBASE:FBan0015238"

/db_xref="FLYBASE:FBgn0029677"

/evidence=not_experimental

/protein_id="AAF45895.1"

/db_xref="GI:7290440"

/translation="MIVNMPVNGTSGASSPTSNPSSNNNSSTTNDNTNLPGDY
NFLKRFDDQDLALGDHYEPIGANPKKRHLDDAGDLHFLQLEOPSPNPTNAAP
AQSTSTFINDLFSVLESSLLPVAAYSHCPSPSSAPLDDNGWGGLLELDHRYNSGLQ

ALQLGNATLPPPPPTSSSAQNASHNTQQTQNVSPGQNPQHNQHFVLPQKQRIQSA
GRYTSVSGANEATEADFEDNLNLLNFKEDFPHLSPHNOIHQGVNLPBGSAP
PFPSTAVPIATSPCSSSPSSASASASAIISTQSLGSLGVLQARSPKSCSSAA
MPAALGGGNGSAGASAAAGGVAGVGSTKTGRKFEELVMEVTSILDGNDMIVAEHVVE
DTSKA"

Join(<12618..12809,12871..>13395)

/gene="CG12632"

/product="CT35179"

/db_xref="FLYBASE:FBan0012632"

/db_xref="FLYBASE:FBgn0029678"

<12618..>13395

/gene="CG12632"

/db_xref="FLYBASE:FBan0012632"

/db_xref="FLYBASE:FBgn0029678"

Join(12618..12809,12871..13395)

/gene="CG12632"

/note="CG12632 gene product"

/codon_start=1

/db_xref="FLYBASE:FBan0012632"

/db_xref="FLYBASE:FBgn0029678"

/protein_id="AAF45896.1"

/db_xref="GI:7290441"

/translation="MLSQHFSDREPYYKSNDRKNSVRHNLSTINPFRKGVKAPOCA
GHLWJSSGDSAEVLAWEHKQKRLDLPFMESINRRIQHQHQQQQLQHQHQHQH
SPQKQSRQAQQHMQPQSSCMYDEAAVAVALITQEMQTNSSGGSTETSOQQQQQQ
HHHSHQNHQNHQRLPFDLLSDDELKRTKQAILNGIHREVEVQSVNSIIST
YHVDLLDNGELTLILFILIP"

Join(<31549..31617,31706..32372,32608..33318,33401..33701,
33926..>34127)

/gene="CG2901"

/product="CT9880"

/db_xref="FLYBASE:FBan0002901"

/db_xref="FLYBASE:FBgn0029679"

/evidence=not_experimental

<31549..>34127

/gene="CG2901"

/db_xref="FLYBASE:FBan0002901"

/db_xref="FLYBASE:FBgn0029679"

/evidence=not_experimental

Join(31549..31617,31706..32372,32608..33318,33401..33701,
33926..34127)

/gene="CG2901"

/note="CG2901 gene product"

/codon_start=1

/db_xref="FLYBASE:FBan0002901"

/db_xref="FLYBASE:FBgn0029679"

/evidence=not_experimental

/protein_id="AAF45897.1"

/db_xref="GI:7290442"

/translation="MKFGKTYESHLTIEWRQYMYGDLKELIKQGVENAPSLTSSD
YEQAYYKAFEEFTFCQSELFGVNNFFLEKLEARRKHGLQLLAYSRBPQHTG
SDSLSORASOKLMTQRLQRYAEFVLSLVLTQNTQSLNETFRKICKYDKNM
RSVAGRVFVNDLPDTRVLRQMTLEVEDLYTHLANGDSRLAMEKLRVPLGE
PTPSPVFRAGIALMLMLLVATAISYWKRAPLEDHTGLMLRFLGPTFWIFNRYM
AANVAGQAGVNHILIFEIDPRSHQPATFLEACTFGILWALSMLGFLYNDLIGVY
DPYVFLGLIMVLVPLPIMNPARWFTIKLVGRVITAPLHYGVFADFWMGDQM
NSLVCIDHYITVRYFAISWLRDVRNCCFEPDVMPVITMCLPGVFRFAQCLRRFD
SLGKSMYILNACKYTTFLVFLSTLRRNSGGYANTFNSPTWFLSSCVVATYIC
YLWDVRLDGLFPRIMRGERIFLRKQLVYQAFYFVIVENLRLFWAVEFTILYHNL
MTPYNMRTISSILEITRRITWNYVRLNEHFNCGNFRATRDHLAALNPRQRMLES
MDESQGVSNRRKSNRIRLGRKF"

Join(<53534..53962,54226..54507,55188..55316,
55391..>59836)

/gene="EG:EG0002.3"

/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."

/product="CT9906"

/db_xref="FLYBASE:FBan0002904"

/db_xref="FLYBASE:FBgn0025376"

<53534..>59836

/gene="EG:EG0002.3"

/note="CG2904"

/map="3F-3F"

```

/db_xref="FLYBASE:FBan0002904"
/db_xref="FLYBASE:FBgn0025376"
join(5334..5396,54226..54507,55188..55316,55391..59836)
/genes="EG:EG00002.3"
/notes="EG:EG00002.3 gene product: Nucleotide sequence of
the Celera sequence differs from the published sequence
for this transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0002904"
/db_xref="FLYBASE:FBgn0025376"
/protein_id="AAF45898.1"
/db_xref="GI:7290443"
/translation="MIKRYKKEPTNVGVAAPTAGAAALGSGASTPGTSPTSPVS
VPVVOAHPHOLGHLISNNGTIPRSRPRKLILDGADKATLRQSKQKXSALA
KVASSVELAYVINGTNGNSHNSAGNSRDCINAVIQLTSSSPALCPPELRR
KVASGLAGRRPPLCGDAAECFELLHRVHSHISPDGDCSSACIAHRRFAMRY
IEQVCYKCGANSEQLFTQMVIYVSASALTSQKSLALSHQQLSFGQLLRAGNMGI

alignment_scores:
    Quality: 10.00      Length: 10
    Ratio: 1.000       Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x AE003429/rev ..

Align seg 1/1 to reverse of: AE003429 from: 1 to: 297405

227 ProProHisHisHisHisHisPro 236
|||||
229639 CCCCCCCCCATCACCACCACCACCACC 229610

seq_name: gb_pat:A57962

seq_documentation_block:
LOCUS A57962 54 bp DNA PAT 05-MAR-1998
DEFINITION Sequence 28 from Patent EP0743364.
ACCESSION A57962
VERSION A57962.1 GI:3713732
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Narwa,R. and Roques,P.
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding
fragments and their application as reactives for risk evaluation of
HIV-1 mother-foetal transmission
JOURNAL Patent: EP 0743364-A 28 20-NOV-1996;
COMMISSARIAT ENERGIE ATOMIQUE (FR)
COMMENT Other publication FR 2734281 961122.
FEATURES
source
location/Qualifiers
1..54
/organism="unidentified"
/db_xref="taxon:32644"
6 t

BASE COUNT 29 a 6 c 13 g
ORIGIN

alignment_scores:
    Quality: 9.00      Length: 9
    Ratio: 1.000       Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x A57962/rev ..

Align seg 1/1 to reverse of: A57962 from: 1 to: 54

972 LeuPheSerThrSerArgAlaSer 980
|||||
29 CTCCTTTCCACTTCTCTAGAGCTTCT 3

seq_name: gb_pat:A57963

seq_documentation_block:
LOCUS MUSIGHYA2 224 bp DNA ROD 20-MAR-2000
DEFINITION Mouse Ig heavy-chain variable region (V) pseudogene V104A DNA, 5'
flanking region.
ACCESSION M17576
VERSION M17576.1 GI:196232
KEYWORDS V-region; immunoglobulin heavy chain; pseudogene.
SEGMENT 2 of 5
SOURCE Mouse DNA from hybridoma B1-8 library.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 224)
AUTHORS Blankenstein,T., Bonhomme,F. and Krawinkel,U.
TITLE Evolution of pseudogenes in the immunoglobulin VH-gene family of
the mouse
JOURNAL Immunogenetics 26, 237-248 (1987)
MEDLINE
FEATURES
source
location/Qualifiers
1..224
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="chi04"
/clone_lib="charon 4A"
61 t

BASE COUNT 70 a 52 c 41 g
ORIGIN About 167 bp downstream of segment 1.

alignment_scores:
    Quality: 9.00      Length: 9
    Ratio: 1.000       Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x A57962/rev ..

Align seg 1/1 to reverse of: A57962 from: 1 to: 54

972 LeuPheSerThrSerArgAlaSer 980
|||||
29 CTCCTTTCCACTTCTCTAGAGCTTCT 3

```

```

seq_name: gb_pat:A57963

seq_documentation_block:
LOCUS A57963 54 bp DNA PAT 05-MAR-1998
DEFINITION Sequence 29 from Patent EP0743364.
ACCESSION A57963
VERSION A57963.1 GI:3713733
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Narwa,R. and Roques,P.
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding
fragments and their application as reactives for risk evaluation of
HIV-1 mother-foetal transmission
JOURNAL Patent: EP 0743364-A 29 20-NOV-1996;
COMMISSARIAT ENERGIE ATOMIQUE (FR)
COMMENT Other publication FR 2734281 961122.
FEATURES
source
location/Qualifiers
1..54
/organism="unidentified"
/db_xref="taxon:32644"
6 t

BASE COUNT 30 a 5 c 13 g
ORIGIN

alignment_scores:
    Quality: 9.00      Length: 9
    Ratio: 1.000       Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x A57963/rev ..

Align seg 1/1 to reverse of: A57963 from: 1 to: 54

972 LeuPheSerThrSerArgAlaSer 980
|||||
29 CTCCTTTCCACTTCTCTAGAGCTTCT 3

seq_name: gb_ro:MUSIGHYA2

seq_documentation_block:
LOCUS MUSIGHYA2 224 bp DNA ROD 20-MAR-2000
DEFINITION Mouse Ig heavy-chain variable region (V) pseudogene V104A DNA, 5'
flanking region.
ACCESSION M17576
VERSION M17576.1 GI:196232
KEYWORDS V-region; immunoglobulin heavy chain; pseudogene.
SEGMENT 2 of 5
SOURCE Mouse DNA from hybridoma B1-8 library.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 224)
AUTHORS Blankenstein,T., Bonhomme,F. and Krawinkel,U.
TITLE Evolution of pseudogenes in the immunoglobulin VH-gene family of
the mouse
JOURNAL Immunogenetics 26, 237-248 (1987)
MEDLINE
FEATURES
source
location/Qualifiers
1..224
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="chi04"
/clone_lib="charon 4A"
61 t

BASE COUNT 70 a 52 c 41 g
ORIGIN About 167 bp downstream of segment 1.

alignment_scores:
    Quality: 9.00      Length: 9
    Ratio: 1.000       Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x A57962/rev ..

Align seg 1/1 to reverse of: A57962 from: 1 to: 54

972 LeuPheSerThrSerArgAlaSer 980
|||||
29 CTCCTTTCCACTTCTCTAGAGCTTCT 3

```

```

Ratio: 1.000          Caps: 0
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x MUSIGHYA2  ..
Align seg 1/1 to: MUSIGHYA2  from: 1 to: 224

227 ProProHishHishHishHis 235
4 CCTCTCCACATCATCATCATCAT 30

seq_name: gb_vil:HIV1GNC

seq_documentation_block:
LOCUS      HIV1GNC      387 bp      DNA      29-AUG-1996
DEFINITION Human immunodeficiency virus type 1 gene coding for matrix protein
P17, partial, isolate IGN-C.
ACCESSION  Z79576.1 GI:1515242
VERSION    Z79576.1
KEYWORDS   matrix protein P17.
SOURCE     Human immunodeficiency virus type 1.
ORGANISM   Human immunodeficiency virus type 1.
REFERENCE  1 (bases 1 to 387)
AUTHORS    Narwa,R., Roques,P., Courpotin,C., Parnet-Mathieu,F., Boussin,F.,
            Roane,A., Marce,D., Lasfargues,G. and Dormont,D.
TITLE      Characterization of human immunodeficiency virus type 1 p17 matrix
            protein motifs associated with mother-to-child transmission
JOURNAL    J. Virol. 70 (7), 4474-4483 (1996)
MEDLINE    96256758
REFERENCE  2 (bases 1 to 387)
AUTHORS    Roques,P.
TITLE      Direct Submission
JOURNAL    Submitted (22-AUG-1996) P. Roques, CEA, Departement de Recherche
            Medicale, 60-68, Av. Division Leclerc, BP 6, F92265 Fontenay Aux
            Roses, FRANCE
FEATURES   Location/Qualifiers
            source
            1..387
            /organism="Human immunodeficiency virus type 1"
            /isolate="IGN-C"
            /db_xref="taxon:11676"
            1..>387
            /codon_start=1
            /product="matrix protein P17"
            /protein_id="CA01829.1"
            /db_xref="GI:1515243"
            /translation="MGARASILSGGLDWERIRLRPGKKYKKHLYWASREMERF
            AVNPGLETAEGCQAMGQIPALQITELRLSYNTIATIIYCVXQKIEVDTKEALE
            EVEKQKQKIQQAARDEGNNSQSNF"
BASE COUNT 156 a 56 c 101 g 63 t 1 others
ORIGIN

alignment_scores:
Quality: 9.00          Length: 9
Ratio: 1.000          Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x HIV1GNC/rev  ..
Align seg 1/1 to reverse of: HIV1GNC  from: 1 to: 387

972 LeuPheSerThrSerArgAlaSer 980
|||||
320 CTCCTTCCACTTCTCTAGAGCTTCT 294

seq_name: gb_vil:HIV1DUBC

seq_documentation_block:
LOCUS      HIV1DUBC      396 bp      DNA      29-AUG-1996
DEFINITION Human immunodeficiency virus type 1 gene coding for matrix protein
P17, partial, isolate DUB-C.
ACCESSION  Z79575.1 GI:1515222
VERSION    Z79575.1
KEYWORDS   matrix protein P17.
SOURCE     Human immunodeficiency virus type 1.
ORGANISM   Human immunodeficiency virus type 1.
REFERENCE  1 (bases 1 to 396)
AUTHORS    Narwa,R., Roques,P., Courpotin,C., Parnet-Mathieu,F., Boussin,F.,
            Roane,A., Marce,D., Lasfargues,G. and Dormont,D.
TITLE      Characterization of human immunodeficiency virus type 1 p17 matrix
            protein motifs associated with mother-to-child transmission
JOURNAL    J. Virol. 70 (7), 4474-4483 (1996)
MEDLINE    96256758
REFERENCE  2 (bases 1 to 396)
AUTHORS    Roques,P.
TITLE      Direct Submission
JOURNAL    Submitted (22-AUG-1996) P. Roques, CEA, Departement de Recherche
            Medicale, 60-68, Av. Division Leclerc, BP 6, F92265 Fontenay Aux
            Roses, FRANCE
FEATURES   Location/Qualifiers
            source
            1..396
            /organism="Human immunodeficiency virus type 1"
            /isolate="DUB-C"
            /db_xref="taxon:11676"
            1..>396
            /codon_start=1
            /product="matrix protein P17"
            /protein_id="CA01828.1"
            /db_xref="GI:1515223"
            /translation="MGARASVLGGELDRWEKIRLRPGKKYKYLKHIWASRELERE
            AVNPGLLTSSGCRQIMGQLPSLQITGTEELKSLYNTVTVLVCVXQKIDVDTKEALE
            EVEKQKISQKIQQAADKNSRVSQNY"
BASE COUNT 165 a 63 c 94 g 73 t 1 others
ORIGIN

alignment_scores:
Quality: 9.00          Length: 9
Ratio: 1.000          Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x HIV1DUBC/rev  ..
Align seg 1/1 to reverse of: HIV1DUBC  from: 1 to: 396

972 LeuPheSerThrSerArgAlaSer 980
|||||
320 CTCCTTCCACTTCTCTAGAGCTTCT 294

seq_name: gb_vil:HIV1MPAC

seq_documentation_block:
LOCUS      HIV1MPAC      396 bp      DNA      29-AUG-1996
DEFINITION Human immunodeficiency virus type 1 gene coding for matrix protein
P17, partial, isolate NPA-C.
ACCESSION  Z79574.1 GI:1515264
VERSION    Z79574.1
KEYWORDS   matrix protein P17.
SOURCE     Human immunodeficiency virus type 1.
ORGANISM   Human immunodeficiency virus type 1.
REFERENCE  1 (bases 1 to 396)
AUTHORS    Narwa,R., Roques,P., Courpotin,C., Parnet-Mathieu,F., Boussin,F.,
            Roane,A., Marce,D., Lasfargues,G. and Dormont,D.
TITLE      Characterization of human immunodeficiency virus type 1 p17 matrix
            protein motifs associated with mother-to-child transmission

```

```

J. Virol. 70 (7), 4474-4483 (1996)
MEDLINE
REFERENCE 2 (bases 1 to 396)
AUTHORS Roques,P.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1996) P. Roques, CEA, Departement de Recherche
Medicale, 60-68, Av. Division Leclerc, BP 6, F92365 Fontenay Aux
Roses, FRANCE
FEATURES
    source      Location/Qualifiers
    1..396
    /organism="Human immunodeficiency virus type 1"
    /isolate="MPA-C"
    /db_xref="taxon:11676"
    1..>396
    /codon_start=1
    /product="matrix protein p17"
    /protein_id="CA01827.1"
    /db_xref="GI:1515265"
    /db_xref="SPTREMBL:O97721"
    /translation="MGARASVLGGKLDFAEKIRLRPGGKKYKLVHIVASRELRF
    ALNPGLETTGCGQIIRQLQPSLQGTGTELSLYNTVTVLYCVHQRIDVRDTKEALE
    EVEKRNKISQOKIQQAADKENSROYSONY"
BASE COUNT 170 a 62 c 90 g 74 t
ORIGIN

alignment_scores:
    Quality: 9.00 Length: 9
    Ratio: 1.000 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
    US-09-332-522b-2 x HIVIMPAC/rev ..
    Align seg 1/1 to reverse of: HIVIMPAC from: 1 to: 396

    972 LeuPheSerThrSerArgAlaSer 980
    |||||
    320 CTCCTTTCCACTTCCTCTAGAGCTTCT 294

seq_name: gb_vil:HIVIMPAC

seq_documentation_block:
    LOCUS HIVIMPAC 396 bp DNA VRL 29-AUG-1996
    DEFINITION Human immunodeficiency virus type 1 gene coding for matrix protein
    P17, partial, isolate MPA-M.
    ACCESSION 279573
    VERSION 279573.1 GI:1515266
    KEYWORDS matrix protein p17
    SOURCE Human immunodeficiency virus type 1.
    ORGANISM Human immunodeficiency virus type 1.
    Viruses; Retroviral viruses; Retroviridae; Lentivirus; Primate
    lentivirus group.
    REFERENCE 1 (bases 1 to 396)
    AUTHORS Narva,R., Roques,P., Courpottin,C., Parnet-Mathieu,F., Boussin,F.,
    Roane,A., Marce,D., Lasfargues,G. and Dormont,D.
    TITLE Characterization of human immunodeficiency virus type 1 p17 matrix
    protein motifs associated with mother-to-child transmission
    J. Virol. 70 (7), 4474-4483 (1996)
    MEDLINE
    REFERENCE 2 (bases 1 to 396)
    AUTHORS Roques,P.
    TITLE Direct Submission
    JOURNAL Submitted (22-AUG-1996) P. Roques, CEA, Departement de Recherche
    Medicale, 60-68, Av. Division Leclerc, BP 6, F92365 Fontenay Aux
    Roses, FRANCE
    FEATURES
        source      Location/Qualifiers
        1..396
        /organism="Human immunodeficiency virus type 1"
        /isolate="MPA-M"
        /db_xref="taxon:11676"
        1..>396
        /codon_start=1

```

```

/product="matrix protein p17"
/protein_id="CA01826.1"
/db_xref="GI:1515267"
/db_xref="SPTREMBL:O97722"
/translation="MGARASVLGGKLDFAEKIRLRPGGKKYKLVHIVASRELRF
ALNPGLETTGCGQIIRQLQPSLQGTGTELSLYNTVTVLYCVHQRIDVRDTKEALE
EVEKRNKISQOKIQQAADKENSROYSONY"
BASE COUNT 165 a 62 c 92 g 76 t
ORIGIN

alignment_scores:
    Quality: 9.00 Length: 9
    Ratio: 1.000 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
    US-09-332-522b-2 x HIVIMPAC/rev ..
    Align seg 1/1 to reverse of: HIVIMPAC from: 1 to: 396

    972 LeuPheSerThrSerArgAlaSer 980
    |||||
    320 CTCCTTTCCACTTCCTCTAGAGCTTCT 294

seq_name: gb_pr3:AF039150

seq_documentation_block:
    LOCUS AF039150 734 bp DNA PRI 02-JUN-1998
    DEFINITION Homo sapiens chromosome 4 cosmid ct379 subclone 1, D424 repeat
    sequence.
    ACCESSION AF039150
    VERSION AF039150.1 GI:3170144
    KEYWORDS human.
    SOURCE Homo sapiens
    ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
    Eutheria; Primates; Catarrhini; Hominidae; Homo.
    REFERENCE 1 (bases 1 to 734)
    AUTHORS Dickson,M.C., Heather,L.J., Bolland,D.J., van Geel,M., de Jong,P.,
    Flint,J., Frants,R.R. and Hewitt,J.E.
    TITLE Inter- and intrachromosomal duplications of human chromosome 4q35:
    an example of telomere plasticity
    JOURNAL Unpublished
    REFERENCE 2 (bases 1 to 734)
    AUTHORS Dickson,M.C.
    TITLE Direct Submission
    JOURNAL Submitted (18-DEC-1997) School of Biological Sciences, 3.239
    Stopford Building, University of Manchester, Oxford Road,
    Manchester M13 9PT, United Kingdom
    FEATURES
        Location/Qualifiers
        1..734
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="4"
        /map="4q35"
        /clone="cosmid ct379 derived from YAC RM2136"
        /sub_clone="1"
        /note="BamHI fragment"
        misc_feature 1..410
        /note="similar to GenBank Accession Number X06587; hhsp3"
        repeat_region <1..>734
        /rpt_type=tandem
        /rpt_family="D424"
        misc_feature 477..734
        /note="double homeobox"
BASE COUNT 107 a 308 c 225 g 94 t
ORIGIN

alignment_scores:
    Quality: 9.00 Length: 9
    Ratio: 1.000 Gaps: 0

```

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x AF039150 ..

Align seg 1/1 to: AF039150 from: 1 to: 734

227 ProProHishHisHishHis 235
 |||||
 128 CCCCCCCCCCACCACCACCACCAC 154

seq_name: gb_pr3:AF039151

seq_documentation_block:

LOCUS AF039151 734 bp DNA PRI 02-JUN-1998
 DEFINITION Homo sapiens chromosome 4 cosmid ct379 subclone 2, D424 repeat
 sequence.

ACCESSION AF039151

VERSION AF039151.1 GI:3170145

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 734)
 Dickson, M.C., Heather, L.J., Bolland, D.J., van Geel, M., de Jong, P.,
 Flint, J., Frants, R.R. and Hewitt, J.E.

TITLE

Inter- and intrachromosomal duplications of human chromosome 4q35:

an example of telomere plasticity

JOURNAL

REFERENCE 2 (bases 1 to 734)

AUTHORS Dickson, M.C.

TITLE

Direct Submission

Submitted (18-DEC-1997) School of Biological Sciences, 3.239

Stopford Building, University of Manchester, Oxford Road,

Manchester M13 9PT, United Kingdom

FEATURES

source

1..734
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4q35"
 /clone="cosmid ct379 derived from YAC 25C2E"
 /sub_clone="2"
 /note="BamHI fragment"

misc_feature

1..410
 /note="similar to GenBank Accession Number X06587; hhspm3"

repeat_region

<1..>734

/rpt_type=tandem

/rpt_family="D424"

misc_feature

477..734

/note="partial double homeobox"

BASE COUNT

107 a 308 c 225 g 94 t

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x AF039151 ..

Align seg 1/1 to: AF039151 from: 1 to: 734

227 ProProHishHisHishHis 235
 |||||
 128 CCCCCCCCCCACCACCACCACCAC 154

seq_name: gb_pr3:AF039152

seq_documentation_block:

LOCUS AF039152 735 bp DNA PRI 02-JUN-1998
 DEFINITION Homo sapiens chromosome 4 cosmid ct379 subclone 3, D424 repeat
 sequence.

ACCESSION AF039152

VERSION AF039152.1 GI:3170146

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 735)
 Dickson, M.C., Heather, L.J., Bolland, D.J., van Geel, M., de Jong, P.,
 Flint, J., Frants, R.R. and Hewitt, J.E.

TITLE

Inter- and intrachromosomal duplications of human chromosome 4q35:

an example of telomere plasticity

JOURNAL

REFERENCE 2 (bases 1 to 735)

AUTHORS Dickson, M.C.

TITLE

Direct Submission

Submitted (18-DEC-1997) School of Biological Sciences, 3.239

Stopford Building, University of Manchester, Oxford Road,

Manchester M13 9PT, United Kingdom

FEATURES

source

1..735
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4q35"
 /clone="cosmid ct379 derived from YAC 25C2E"
 /sub_clone="3"
 /note="BamHI fragment"

misc_feature

1..411

/note="similar to GenBank Accession Number X06587; hhspm3"

repeat_region

<1..>735

/rpt_type=tandem

/rpt_family="D424"

misc_feature

478..735

/note="partial double homeobox"

BASE COUNT

108 a 307 c 227 g 93 t

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x AF039152 ..

Align seg 1/1 to: AF039152 from: 1 to: 735

227 ProProHishHisHishHis 235
 |||||
 129 CCCCCCCCCCACCACCACCACCAC 155

seq_name: gb_pr3:AF039154

seq_documentation_block:

LOCUS AF039154 735 bp DNA PRI 02-JUN-1998

DEFINITION Homo sapiens chromosome 4 cosmid B2 subclone 2, D424 repeat
 sequence.

ACCESSION AF039154

VERSION AF039154.1 GI:3170148

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 735)
 Dickson, M.C., Heather, L.J., Bolland, D.J., van Geel, M., de Jong, P.,
 Flint, J., Frants, R.R. and Hewitt, J.E.

TITLE

Inter- and intrachromosomal duplications of human chromosome 4q35:

```

an example of telomere plasticity
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 735)
AUTHORS Dickson, M.C.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-1997) School of Biological Sciences, 3.239
Stopford Building, University of Manchester, Oxford Road,
Manchester M13 9PT, United Kingdom
FEATURES
source
Location/Qualifiers
1..735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4q35"
/clone="cosmid B2 derived from YAC RM2173"
/sub_clone="2"
/notes="BamHI fragment"
misc_feature 1..411
/notes="similar to GenBank Accession Number X06587; hhspm3"
repeat_region <1..>735
/rpt_type=tandem
/rpt_family="D424"
misc_feature 478..735
/notes="partial double homeobox"
BASE COUNT 109 a 306 c 225 g 95 t
ORIGIN
alignment_scores
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block
US-09-332-522B-2 x AF039154 ..
Align seg 1/1 to: AF039154 from: 1 to: 735
227 ProProHSHSHSHSHSHSHSHSHS 235
|||||
129 CCCCCCCCCCACCACCACCACCAC 155
seq_name: gb_pr3:AF039155
seq_documentation_block:
LOCUS AF039155 735 bp DNA PRI 02-JUN-1998
DEFINITION Homo sapiens chromosome 4 cosmid B2 subclone 3, D424 repeat
sequence.
ACCESSION AF039155
VERSION AF039155.1 GI:3170149
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 735)
AUTHORS Dickson, M.C., Heather, L.J., Bolland, D.J., van Geel, M., de Jong, P.,
Flint, J., Frants, R.R. and Hewitt, J.E.
TITLE Inter- and intrachromosomal duplications of human chromosome 4q35:
an example of telomere plasticity
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 735)
AUTHORS Dickson, M.C.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-1997) School of Biological Sciences, 3.239
Stopford Building, University of Manchester, Oxford Road,
Manchester M13 9PT, United Kingdom
FEATURES
source
Location/Qualifiers
1..735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4q35"

```

```

/clone="cosmid B2 derived from YAC RM2173"
/sub_clone="3"
/notes="BamHI fragment"
misc_feature 1..411
/notes="similar to GenBank Accession Number X06587; hhspm3"
repeat_region <1..>735
/rpt_type=tandem
/rpt_family="D424"
misc_feature 478..735
/notes="partial double homeobox"
BASE COUNT 108 a 308 c 225 g 94 t
ORIGIN
alignment_scores
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block
US-09-332-522B-2 x AF039155 ..
Align seg 1/1 to: AF039155 from: 1 to: 735
227 ProProHSHSHSHSHSHSHSHSHS 235
|||||
129 CCCCCCCCCCACCACCACCACCAC 155
seq_name: gb_om:BOVVEGFB
seq_documentation_block:
LOCUS BOVVEGFB 802 bp mRNA MAM 27-APR-1993
DEFINITION Bovine vascular endothelial growth factor (VEGF-B) mRNA, 3' end.
ACCESSION M33750
VERSION M33750.1 GI:163810
KEYWORDS vascular endothelial growth factor.
SOURCE Bovine, CDNA to mRNA, clone lambda-ST800.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Bovinae; Bos.
REFERENCE 1 (bases 1 to 802)
AUTHORS Tischer, E., Gospodarowicz, D., Mitchell, R.W., Silva, M.,
Schilling, J., Lau, K., Crisp, T.M., Fiddes, J.C. and Abraham, J.A.
TITLE Vascular endothelial growth factor: A new member of the
platelet-derived growth factor gene family
JOURNAL Biochem. Biophys. Res. Commun. 165, 1198-1206 (1989)
MEDLINE 90121225
FEATURES
source
Location/Qualifiers
1..802
/organism="Bos taurus"
/db_xref="taxon:9913"
<1..802
/notes="vascular endothelial growth factor mRNA"
/codon_start=1
/protein_id="AAA30805.1"
/db_xref="GI:163811"
/translation="APMAGGQKPEHVKFMVYQSFPCPIETLYDIFQYDEIEF
IFPKSVPLMRGCGCCNDELCVPTFEETNITQIMRKIPHQSHIGENSFLQHNKE
CRPKDKARQEKDKPRR"
BASE COUNT 241 a 195 c 189 g 177 t
ORIGIN
alignment_scores
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block
US-09-332-522B-2 x BOVVEGFB ..

```

Align seg 1/1 to: BOVVEGFB from: 1 to: 802

227 ProProHisHisHisHisHisHis 235
|||||
456 CCACCACCACCACCACCACCACCAT 482

seq_name: gb_om:BOVVEGFA

seq_documentation_block:
LOCUS BOVVEGFA 934 bp mRNA MAM 27-APR-1993
DEFINITION Bovine vascular endothelial growth factor (VEGF-A) mRNA, 3' end.
ACCESSION M31836
VERSION M31836.1 GI:163808
KEYWORDS vascular endothelial growth factor.
SOURCE Bovine, cDNA to mRNA, clone lambda-Sr800.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Bovinae; Bos.

REFERENCE
AUTHORS Fischer, E., Gospodarowicz, D., Mitchell, R.W., Silva, M.,
Schilling, J., Lau, K., Crisp, T.M., Fiddes, J.C. and Abraham, J.A.
TITLE Vascular endothelial growth factor: A new member of the
platelet-derived growth factor gene family
JOURNAL Biochem. Biophys. Res. Commun. 165, 1198-1206 (1989)
MEDLINE 90121225

FEATURES
source Location/Qualifiers
1..934
/organism="Bos taurus"
/db_xref="taxon:9913"
<1..934
/note="vascular endothelial growth factor mRNA"
CDS
<1..495
/note="vascular endothelial growth factor"
/codon_start=1
/protein_id="BAA30804.1"
/db_xref="GI:163809"

BASE COUNT 277 a 225 c 225 g 207 t
ORIGIN
|||||
KPRR

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x BOVVEGFA ..

Align seg 1/1 to: BOVVEGFA from: 1 to: 934

227 ProProHisHisHisHisHisHis 235
|||||
588 CCACCACCACCACCACCACCACCAT 614

seq_name: gb_v11:AB021155

seq_documentation_block:
LOCUS AB021155 975 bp DNA VRL 06-AUG-1999
DEFINITION Unidentified baculovirus HHNBV-XIA gene, complete cds.
ACCESSION AB021155
VERSION AB021155.1 GI:5706373
KEYWORDS HHNBV-XIA
SOURCE unidentified baculovirus (sub_species:hypodermal and hematopoietic
necrosis baculovirus) DNA.
ORGANISM unidentified baculovirus
VIRUSES; dsDNA viruses, no RNA stage; Baculoviridae.
REFERENCE 1 (bases 1 to 975)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Xia, C. and Liu, J.
Hypodermal and hematopoietic necrosis baculovirus
Published Only in Database (1999) In press
2 (bases 1 to 975)
Xia, C. and Liu, J.
Direct Submission
Submitted (13-DEC-1998) to the DDBJ/EMBL/GenBank databases. Chun
Xia, College of Veterinary Medicine, China Agricultural University,
Microbiology and Infection; Yuan Ming Yuan 2, Beijing 100094, P. R.
China (E-mail:xiachun@public.east.cn.net, Tel:86-10-62893154)

FEATURES
source Location/Qualifiers
1..975

/organism="unidentified baculovirus"
/sub_species="hypodermal and hematopoietic necrosis
baculovirus"
/db_xref="taxon:10469"
complement(185..832)
/gene="HHNBV-XIA"
complement(185..832)
/gene="HHNBV-XIA"
/codon_start=1
/product="HHNBV-XIA"
/protein_id="BAA83091.1"
/db_xref="GI:5706374"
/translation="MCNVSLVGYKLSDNAIQYHESLCSLHSSISYADYYTSRNN
SEGGNSSSEKSNADVAKTMASFYDFKSDSKNNKNTSNEILIKMFQMDRVLDG
MDDDEDSSSENEEEEEIVKPKARKVEDVDSNKTLPKPAVKYKQVED
VEMEVEKAAAEKEEKEQAKEDATEYDDDEDEKAVASDEDEDESKAIF"

BASE COUNT 236 a 238 c 138 g 363 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x AB021155/rev ..

Align seg 1/1 to reverse of: AB021155 from: 1 to: 975

327 AspGluAspGluAspGluaspSer 335
|||||
226 GATGAAGATGAAGATGATGAAGATTCT 200

seq_name: gb_pat:A32895

seq_documentation_block:
LOCUS A32895 1282 bp mRNA PAT 19-JUL-1996
DEFINITION R-prolixus thrombin inhibitor protein coding cDNA.
ACCESSION A32895
VERSION A32895.1 GI:1567709

KEYWORDS Rhodnius prolixus.
SOURCE Rhodnius prolixus.
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Hemiptera; Euhemiptera; Heteroptera; Panheteroptera;
Cimicomorpha; Reduviidae; Triatominae; Rhodnius.

REFERENCE
AUTHORS
TITLE NOVEL THROMBIN INHIBITOR PROTEIN FROM REDUVII
JOURNAL Patent: WO 9309232-A 16 13-MAY-1993;
FEATURES Location/Qualifiers
source 1..1282
/organism="Rhodnius prolixus"
/db_xref="taxon:13249"
<1..1109
/codon_start=3
/product="thrombin inhibitor protein"
/protein_id="CAA02124.1"
/db_xref="GI:1567710"

/translation="RLLLLGLAALVAEGEGPCACPHALHRVCGSDGETYSNPCTLN

CAKFKGKPELVKHDPCEPDEDVCOECGDDEVKPVCGSDDDITYDNNCRLECASIS
SSPGVELKHGECRTEEEKILKRSEDEFMYRCACPKIYYPVCGTGGTYPNLCVLECH
MMNPGQLHHYGHCOHHHHPPPHHHHHHTTEKPEPCACPHALHRVCGSDG
ETYSNPCTLNCAKHGKPLVKVHDGPCEPDEDVCOECDDVDYEPVCGTDDKTYDN
NCRLECASISSPGVELKHGECRTEEEKILKRSEGLBHTCVCYNYFPVCGTGGTY
PNLCALQCRMREVPCKLKHGKPLHLDFDPV"
BASE COUNT 416 a 261 c 268 g 337 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x A32895 ..

Align seg 1/1 to: A32895 from: 1 to: 1282

227 ProProHisHisHisHisHisHis 235
|||||
549 CCTCCTCCTCATCATCATCATCAT 575

seq_name: gb_pat:121698

seq_documentation_block: 1282 bp DNA PAT 07-OCT-1996
LOCUS I21698
DEFINITION Sequence 16 from patent US 5523287.
ACCESSION I21698
VERSION I21698.1 GI:1602052
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1282)
AUTHORS Friedrich,T., Bialojan,S., Kroeger,B. and Kuenast,C.
TITLE Thrombin-inhibitory protein from assassin bugs
JOURNAL Patent: US 5523287-A 16 04-JUN-1996;
FEATURES
Location/Qualifiers
1..1282
/organism="unknown"
BASE COUNT 416 a 261 c 268 g 337 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x I21698 ..

Align seg 1/1 to: I21698 from: 1 to: 1282

227 ProProHisHisHisHisHisHis 235
|||||
549 CCTCCTCCTCATCATCATCATCAT 575

seq_name: gb_v11:MWI012282

seq_documentation_block: 1405 bp DNA VRL 22-JAN-1999
LOCUS MWI012282
DEFINITION Myxoma virus mt3, t3c and mt4 (partial) genes.
ACCESSION AJ012282
VERSION AJ012282.1 GI:4186088
KEYWORDS
mt3 gene; MT3 protein; mt4 gene; MT4 protein; t3c gene; t3c
protein.

SOURCE Myxoma virus.
ORGANISM Myxoma virus
Viruses: dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
Leporipoxvirus.

REFERENCE 1 (bases 1 to 1405)
AUTHORS Becher,D.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-1998) Becher D., Biotechnikum Greifswald,
Micromun GmbH, Walther Rathenau Strasse 49A, Greifswald, D-17489
Greifswald, GERMANY

REFERENCE 2 (bases 1 to 1405)

AUTHORS Salim,S., Meissner,K., Neubert,A., Doehner,I. and Becher,D.
TITLE The MT3 gene locus region of Myxoma virus contains several open
reading frames

JOURNAL Unpublished

FEATURES
source

Location/Qualifiers
1..1405
/organism="Myxoma virus"
/virion

gene /strain="Lausanne"

CDS /db_xref="taxon:10273"

1..123

/gene="mt4"

<1..123

/gene="mt4"

/codon_start=1

/product="MT4 protein"

/protein_id="CAA09972.1"

/db_xref="GI:4186089"

/translation="GSOHEHYLQTCDRNVPDRRTFKAYKPHQRPYAKVLRDEL"

355..696

/gene="t3c"

355..696

/gene="t3c"

/codon_start=1

/product="T3c protein"

/protein_id="CAA09973.1"

/db_xref="GI:4186090"

/translation="MLCVFYLRCLNLFVSVYTLVMPINRLVSMFGEINPVCEPD
PKKDDRRVVPADCPPIPKPELNANELNGMDFMKIPNPFRRYECNNNTKOPSSYPLK
KGFDMETLE"

766..1221

/gene="mt3"

766..1221

/gene="mt3"

/note="putative"

/codon_start=1

/product="MT3 protein"

/protein_id="CAA09974.1"

/db_xref="GI:4186091"

/translation="MTFVVRNEFGADSVILAVODYLRWSATPPDRRESAGRLFAVE
ESFKRDAEIVFRDNVRLVRCYSYEHVDSSESLIRMLRHSYITEACALIGLARVAE
TWGDDPIPTSVRIEFLMRLITDVEVPHLSVINIRLQLCSQERRDD"

BASE COUNT 339 a 348 c 317 g 401 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x MWI012282/rev ..

Align seg 1/1 to reverse of: MWI012282 from: 1 to: 1405

453 ArgAsnGluArgLysArgSerArgMet 461
|||||

759 AGGAATGAAGGAACGCTTCGAGATG 733

seq_name: gb_v11:HIVLBV217

seq_documentation_block:

LOCUS HIVLBV217 1462 bp ss-DNA VRL 26-JUL-1993
DEFINITION Human immunodeficiency virus type 1 gag gene, isolate LBV21-7, 5'
end.

```

ACCESSION      L11778
VERSION        L11778.1  GI:305980
KEYWORDS       gag gene.
SOURCE         Human immunodeficiency virus type 1 (individual isolate LBV21-7) blood DNA.
ORGANISM       Human immunodeficiency virus type 1
               Viruses; Retroviruses; Retroviridae; Lentivirus; Primate lentivirus group.
REFERENCE      1 (bases 1 to 1462)
AUTHORS        Louwagie,J.J., McCutchan,F., Brennan,T., Peeters,M., Brennan,T., Sanders-Buell,E., Eddy,G., van der Groen,G., Fransen,K., Gershy-Damet,M., Deleys,R. and Burke,D.
TITLE          Phylogenetic analysis of gag genes from seventy international HIV-1 isolates provides evidence for multiple genotypes
JOURNAL        AIDS 7, 769-780 (1993)
MEDLINE        93371703
FEATURES       Location/Qualifiers
               1..1462
               /organism="Human immunodeficiency virus type 1"
               /proviral
               /isolate="LBV21-7"
               /db_xref="taxon:11676"
               /cell_type="lymphocyte"
               /tissue_type="blood"
               1..1462
               /gene="gag"
               1..1462
               /partial
               /gene="gag"
               /codon_start=1
               /product="gag protein"
               /protein_id="AA44774.1"
               /db_xref="GI:54975"
               /translation="NGARASVLSGGKLDWEKIRLRPGKKYRMKHLVWASRELERF
               ALNPGLEETEGCQIQLQPSLQTGTEELSLFNNAVAVLWCIHQIRIGVKTKEALE
               EVKRNKNSQKIQQADKGNDSQVSNYPVQNAQGMVHOPISTPLNAAVKKVVEEK
               AFSEVTPMSELSQATQDLNMTLVGGGAQOMKLDIINEEAEDRIHQQA
               GPIPPGQIRDPGRSDIAGTSTLQEIIRWNTSNPPIPVGEIKRWILGNLKVYMS
               PVSILDIRQKPFKFDVDFKTLRAEQATQDVANWMTDLLVQNAFNPDCITLKA
               LGPGATLEEMMTACQGVGSPSHKARVLAEMSQAGTATIMWKGNFGRPNKCF
               DCGEGHARNCRAKRGCKGCKEGHMKDCTERQANFLGKINPShRGRPGNFIQN
               RPEPSAPPALSGFGEIAPSPKPEKLYPLTSLKS"
BASE COUNT    530 a 283 c 371 g 278 t
ORIGIN

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-332-522B-2 x HIVLBV217/rev ..
  Align seg 1/1 to reverse of: HIVLBV217 from: 1 to: 1462
  972 LeuPheSerThrSerArgAlaSer 980
  320 CTCCTTTCCACTCTCTAGAGCTTCT 294
  seq_name: gb_om:BOVEGFA

seq_documentation_block:
  LOCUS      BOVEGFA 1536 bp mRNA NAM 27-APR-1993
  DEFINITION Bovine heparin-binding vascular endothelial growth factor (VEGF) mRNA, complete cds.
  ACCESSION  M32976
  VERSION   M32976.1  GI:163006
  KEYWORDS  angiogenic mitogen; vascular endothelial growth factor.
  SOURCE    Bovine folliculostellate cell, cDNA to mRNA.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;

```

```

REFERENCE      Bovinae: Bos.
AUTHORS        1 (bases 1 to 1536)
TITLE          Leung,D.W., Cachianes,G., Kuang,W.-J., Goeddel,D.V. and Ferrara,N.
               Vascular endothelial growth factor is a secreted angiogenic mitogen
JOURNAL        Science 246, 1306-1309 (1989)
MEDLINE        90069608
FEATURES       Location/Qualifiers
               1..1536
               /organism="Bos taurus"
               /db_xref="taxon:9913"
               <1..1536
               /note="VEGF mRNA"
               546..623
               sig_peptide
               /note="vascular endothelial growth factor signal peptide"
               CDS
               546..1118
               /note="vascular endothelial growth factor (VEGF) precursor"
               /codon_start=1
               /protein_id="AAA30502.1"
               /db_xref="GI:163007"
               /translation="MNFLLSWHWSLALLYLHHAKWSQAAPMARGQKPHVVKFMD
               VYQSFPCPIETLVDIFQYDEIEFIEFKPCVPLMRGGCNDSELCVTEPEFNIT
               MOIMRIKPHQSHIGEMSFLOHNKCECPKKDKARQENPCGPGCERRRHLFVQDPQTC
               KCSCKNTDSRCKARQLELNERTCRCDKPRR"
               624..1115
               mat_peptide
               /note="vascular endothelial growth factor"
BASE COUNT    366 a 431 c 468 g 271 t
ORIGIN

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-332-522B-2 x BOVEGFA ..
  Align seg 1/1 to: BOVEGFA from: 1 to: 1536
  227 ProProHisHisHisHisHisHis 235
  1211 CCACCACCACCACCACCACCACCAT 1237
  seq_name: gb_inl:AB025783

seq_documentation_block:
  LOCUS      AB025783 1616 bp mRNA INV 04-APR-2000
  DEFINITION Octopus vulgaris Ov6as mRNA for G protein a subunit s class, complete cds.
  ACCESSION  AB025783
  VERSION   AB025783.1  GI:7416006
  KEYWORDS  G protein a subunit s class; Octopus vulgaris G protein a subunit class s.
  SOURCE    Octopus vulgaris adult eye tissue_lib:hemisected eye cDNA to mRNA.
  ORGANISM  Octopus vulgaris
            Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Octopoda;
            Incirrata; Octopodidae; Octopus.
REFERENCE    1 (sites)
AUTHORS      Iwasa,T., Yanai,T., Nakagawa,M., Kikkawa,S., Obata,S., Usukura,J. and Tsuda,M.
TITLE        G protein alpha subunit genes in octopus photoreceptor cells
JOURNAL      Unpublished (1999)
REFERENCE    2 (bases 1 to 1616)
AUTHORS      Iwasa,T.
TITLE        Direct Submission
JOURNAL      Submitted (03-APR-1999) to the DDBJ/EMBL/GenBank databases. Tatsuo Iwasa, Himeji Institute of Technology; 1475-2 Kanaji, Kamigori-cho, Ako-gun, Hyogo 678-1297, Japan
               (E-mail: iwasat@sci.himeji-tech.ac.jp, Tel:81-7915-8-0195, Fax:81-7915-8-0197)
FEATURES     Location/Qualifiers
               1..1616

```

/organism="Octopus vulgaris"
/db_xref="taxon:6645"
/dev_stage="adult"
/tissue_lib="hemisected eye"
/tissue_type="eye"
6. .11
TATA_signal 68. .72
CAAT_signal 137. .1270
gene /gene="OvGas"
137. .1270
CDS /gene="OvGas"
/codon_start=1
/product="G protein a subunit s class"
/protein_id="BA93639.1"
/db_xref="GI:7416007"
/translation="MCFRTREDDKLKEQNKKEKQLQKQVYRSTHRLLLG
AGSGSTIVKRIHVDGFGPEERKLIKEDIKNRDAILITIGAMSTLNPVEL
DSSLHMHYIOEIASQDDYPAEFYEHTELWRDKVQTCYERSNEYQLDCAQY
FLDRHVIVSADITPTQEDILRCRVLTSGIFETRFIVDKVNFHFDVGGORDERRWI
QCFNDVTAIFVTACSYNLVLRDNTQNLKESLELQSIWNNRWLRSTSVILELNK
ODLAEKVAGKSKITFYADFERYVTPVDAHTPEGDDPEVVRKVFIRDEFRLRSTA
TRGRHYCYPHFICAVDTENIRRVNDCRDIIGRMHLRQIEL"
BASE COUNT 557 a 262 c 354 g 443 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x AB025783/rev ..
Align seg 1/1 to reverse of: AB025783 from: 1 to: 1616

227 ProProHishHishHishHis 235
1483 CCTCTCTCATCATCATCATCATCAT 1457
|||||

seq_name: gb_pat:114451

seq_documentation_block: 1744 bp DNA PAT 26-SEP-1995
LOCUS 114451
DEFINITION Sequence 25 from patent US 5449768.
ACCESSION 114451
VERSION 114451.1 GI:996934
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1744)
AUTHORS Chakraborty,P.R., Dashkevicz,M., Elbrecht,A., Feighner,S.D.,
Liberator,P.A. and Profous-Juchelka,H.
TITLE Elmeria praecox 16S rDNA probes
JOURNAL Patent: US 5449768-A 25 12-SEP-1995;
FEATURES Location/Qualifiers
source 1. .1744
/organism="unknown"

BASE COUNT 438 a 352 c 478 g 476 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x 114451 ..
Align seg 1/1 to: 114451 from: 1 to: 1744

258 ThrLysProThrHisLeuValGluPro 266
|||||
193 ACAAACCAACCCACCTTGTGGAGCCT 219

seq_name: gb_pat:127294

seq_documentation_block: 1744 bp DNA PAT 06-FEB-1997
LOCUS 127294
DEFINITION Sequence 25 from patent US 5563256.
ACCESSION 127294
VERSION 127294.1 GI:1818070
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1744)
AUTHORS Chakraborty,P.R., Dashkevicz,M., Elbrecht,A., Feighner,S.D.,
Liberator,P.A. and Profous-Juchelka,H.
TITLE Elmeria tenella 16S rDNA probes
JOURNAL Patent: US 5563256-A 25 08-OCT-1996;
FEATURES Location/Qualifiers
source 1. .1744
/organism="unknown"

BASE COUNT 438 a 352 c 478 g 476 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x 127294 ..
Align seg 1/1 to: 127294 from: 1 to: 1744

258 ThrLysProThrHisLeuValGluPro 266
|||||
193 ACAAACCAACCCACCTTGTGGAGCCT 219

seq_name: gb_inl:EBU67116

seq_documentation_block: 1744 bp rRNA INV 07-MAY-1997
LOCUS EBU67116
DEFINITION Elmeria brunetti small subunit ribosomal RNA gene, complete
sequence.
ACCESSION U67116
VERSION U67116.1 GI:1813516
KEYWORDS
SOURCE Elmeria brunetti.
ORGANISM Elmeria brunetti.
Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeriida; Elmeriidae;

REFERENCE 1 (bases 1 to 1744)
AUTHORS Barta,J.R., Martin,D.S., Liberator,P.A., Dashkevicz,M.,
Anderson,J.W., Feighner,S.D., Elbrecht,A., Perkins-Barrow,A.,
Jenkins,M.C., Danforth,H.D., Ruff,M.D. and Profous-Juchelka,H.
TITLE Phylogenetic relationships among eight Elmeria species infecting
domestic fowl inferred using complete small subunit ribosomal DNA
sequences
JOURNAL J. Parasitol. 83 (2), 262-271 (1997)
MEDLINE 97259180
REFERENCE 2 (bases 1 to 1744)
AUTHORS Profous-Juchelka,H.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-1996) Parasite Biochemistry, Merck & Co. Inc., PO
Box 2000, Rahway, NJ 07065, USA
FEATURES Location/Qualifiers
source 1. .1744
/organism="Elmeria brunetti"
/strain="Attenuated E. brunetti VA"
/db_xref="taxon:51314"

```

rRNA
1. .1744
/notes="small subunit ribosomal RNA"
BASE COUNT      438 a      352 c      478 g      476 t
ORIGIN

alignment_scores:
  Quality:      9.00      Length:      9
  Ratio:        1.000     Gaps:        0
  Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x EBU67116 ..
Align seg 1/1 to: EBU67116 from: 1 to: 1744

258 ThrLysProThrHisLeuValGluPro 266
|||||
193 ACAAAACCAACCCACCTGTGGAGCCT 219
|||||

seq_name: gb_inl:CFTRG

seq_documentation_block:
LOCUS      CFTRG      1771 bp      DNA      INV      06-MAY-1997
DEFINITION C.fasciculata gene encoding trypanothione reductase.
ACCESSION  Z12618
VERSION    212618.1 GI:6995
KEYWORDS   trypanothione reductase.
SOURCE     Crithidia fasciculata.
ORGANISM   Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Crithidia.
REFERENCE  1 (bases 1 to 1771)
AUTHORS    Aboagye-Kwarteng,T., Smith,K. and Fairlamb,A.H.
TITLE      Molecular characterization of the trypanothione reductase gene from Crithidia fasciculata and Trypanosoma brucei: comparison with other flavoprotein disulphide oxidoreductases with respect to substrate specificity and catalytic mechanism
JOURNAL    Mol. Microbiol. 6 (21), 3089-3099 (1992)
MEDLINE    93086418
REFERENCE  2 (bases 1 to 1771)
AUTHORS    Hunter,W.N., Bailey,S., Habash,J., Harrop,S.J., Helliwell,J.R., Aboagye-Kwarteng,T., Smith,K. and Fairlamb,A.H.
TITLE      Active site of trypanothione reductase. A target for rational drug design
JOURNAL    J. Mol. Biol. 227 (1), 322-333 (1992)
MEDLINE    92395672
REMARK     (sites)
REFERENCE  3 (bases 1 to 1771)
AUTHORS    Fairlamb,A.H.
TITLE      Direct Submission
JOURNAL    Submitted (13-NOV-1991) A.H. Fairlamb, London School of Hygiene a. Tropical Med., Department of Medical Parasitology, Keppel Street, London WC1E 7HT, UK

FEATURES
  Source
    1..1771
    /organism="Crithidia fasciculata"
    /strain="Wild type"
    /isolate="HS6"
    /db_xref="taxon:5656"
    /clone_lib="EMBL3"
    /clone="pKSTR"
    111..11586
    /standard_name="tr"
    /EC_number="1.6.4.8"
    /function="Flavoprotein disulphide oxidoreductase"
    /note="Active-site cysteine disulphide: cys52 and cys57. Active site base: his461. Homodimer. FAD co-enzyme. Cytosolic protein. Similarity to glutathione reductase and dihydrolipoamide dehydrogenase."

CDS
1..1744
/notes="small subunit ribosomal RNA"
BASE COUNT      438 a      352 c      478 g      476 t
ORIGIN

alignment_scores:
  Quality:      9.00      Length:      9
  Ratio:        1.000     Gaps:        0
  Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x EBU67116 ..
Align seg 1/1 to: EBU67116 from: 1 to: 1744

258 ThrLysProThrHisLeuValGluPro 266
|||||
193 ACAAAACCAACCCACCTGTGGAGCCT 219
|||||

seq_name: gb_inl:CFTRG

seq_documentation_block:
LOCUS      CFTRG      1771 bp      DNA      INV      06-MAY-1997
DEFINITION C.fasciculata gene encoding trypanothione reductase.
ACCESSION  Z12618
VERSION    212618.1 GI:6995
KEYWORDS   trypanothione reductase.
SOURCE     Crithidia fasciculata.
ORGANISM   Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Crithidia.
REFERENCE  1 (bases 1 to 1771)
AUTHORS    Aboagye-Kwarteng,T., Smith,K. and Fairlamb,A.H.
TITLE      Molecular characterization of the trypanothione reductase gene from Crithidia fasciculata and Trypanosoma brucei: comparison with other flavoprotein disulphide oxidoreductases with respect to substrate specificity and catalytic mechanism
JOURNAL    Mol. Microbiol. 6 (21), 3089-3099 (1992)
MEDLINE    93086418
REFERENCE  2 (bases 1 to 1771)
AUTHORS    Hunter,W.N., Bailey,S., Habash,J., Harrop,S.J., Helliwell,J.R., Aboagye-Kwarteng,T., Smith,K. and Fairlamb,A.H.
TITLE      Active site of trypanothione reductase. A target for rational drug design
JOURNAL    J. Mol. Biol. 227 (1), 322-333 (1992)
MEDLINE    92395672
REMARK     (sites)
REFERENCE  3 (bases 1 to 1771)
AUTHORS    Fairlamb,A.H.
TITLE      Direct Submission
JOURNAL    Submitted (13-NOV-1991) A.H. Fairlamb, London School of Hygiene a. Tropical Med., Department of Medical Parasitology, Keppel Street, London WC1E 7HT, UK

FEATURES
  Source
    1..1771
    /organism="Crithidia fasciculata"
    /strain="Wild type"
    /isolate="HS6"
    /db_xref="taxon:5656"
    /clone_lib="EMBL3"
    /clone="pKSTR"
    111..11586
    /standard_name="tr"
    /EC_number="1.6.4.8"
    /function="Flavoprotein disulphide oxidoreductase"
    /note="Active-site cysteine disulphide: cys52 and cys57. Active site base: his461. Homodimer. FAD co-enzyme. Cytosolic protein. Similarity to glutathione reductase and dihydrolipoamide dehydrogenase."

CDS
1..1744
/notes="small subunit ribosomal RNA"
BASE COUNT      438 a      352 c      478 g      476 t
ORIGIN

alignment_scores:
  Quality:      9.00      Length:      9
  Ratio:        1.000     Gaps:        0
  Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x CFTRG ..
Align seg 1/1 to: CFTRG from: 1 to: 1771

745 LeuArgArgAlaArgHisValArg 753
|||||
242 CTACGCCGGCTCGCGGCACGCGGT 268
|||||

seq_name: gb_ro:MUSIGHV11

seq_documentation_block:
LOCUS      MUSIGHV11  1823 bp      DNA      ROD      27-APR-1993
DEFINITION Mouse Ig germline H-chain V-region pseudogene VH104A and 5' flank, VH-II.
ACCESSION  J00503.1 GI:196072
VERSION    J00503.1 GI:196072
KEYWORDS   C-region; V-region; germline; immunoglobulin heavy chain; pseudogene; variable region subgroup VH-II.
SEGMENT    1 of 4
SOURCE     Mouse (BALB/c) embryo DNA, from Charon 4A library, clone pCh104 [1],[2].
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 623 to 1372)
AUTHORS    Givol,D., Zakut,R., Efron,K., Rechavi,G., Ram,D. and Cohen,J.B.
TITLE      Diversity of germ-line immunoglobulin V-H genes
JOURNAL    Nature 292, 426-430 (1981)
MEDLINE    81245215
REFERENCE  2 (bases 1 to 622; 1373 to 1823)
AUTHORS    Cohen,J.B., Efron,K., Rechavi,G., Ben-Neriah,Y., Zakut,R. and Givol,D.
TITLE      Simple DNA sequences in homologous flanking regions near immunoglobulin V-H genes: A role in gene interaction?
JOURNAL    Nucleic Acids Res. 10, 3353-3370 (1982)
MEDLINE    82247212
COMMENT     Four mouse embryonic VH genes (108A, 108B, 104 and 111) show diversity among themselves and with the one mature VH gene studied from this subgroup (MPC11). The diversity is pronounced in the three hypervariable segments coding for the CDR residues. Much of the diversity is argued to be time-dependent evolutionary divergence rather than somatic variation. The termination codon at base 1063 indicates that this is a pseudogene.
FEATURES
  Location/Qualifiers
    1..1823
    /organism="Mus musculus"
    /strain="BALB/c"
    /sub_species="domesticus"
    /db_xref="taxon:10090"
    /dev_stage="embryo"

```

```

sig_peptide      /tissue_lib="Charon 4A"
                  join(808..853,938..948)
                  /gene="pseudo-VH104"
CDS              /note="pseudo-VH104 Ig signal pept"
                  join(808..853,938..1242)
                  /partial
                  /gene="pseudo-VH104"
                  /pseudo
                  /codon_start=1
gene             join(808..853,938..1242)
                  /gene="pseudo-VH104"
exon             <808..853
                  /gene="pseudo-VH104"
                  /note="pseudo-VH104 Ig"
                  /pseudo
                  /number=1
Intron           854..937
                  /gene="pseudo-VH104"
                  /number=1
exon             938..1242
                  /gene="pseudo-VH104"
                  /number=2
mat_peptide      949..1242
                  /partial
                  /gene="pseudo-VH104"
                  /pseudo
misc_signal       1243..1281
                  /note="putative recombination recognition sequence;
                  putative"
IDNA              1243..>1823
BASE COUNT       542 a 373 c 312 g 531 t 65 others
ORIGIN           485 bp upstream of Hinfi site on chromosome 12.

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-332-522B-2 x MUSIGHV11 ..
  Align seg 1/1 to: MUSIGHV11 from: 1 to: 1823

227 ProProHisHisHisHisHis 235
|||||
391 CCTCTCCACATCATCATCATCAT 417

seq_name: gb_in1:TRFTRPREDB

seq_documentation_block:
LOCUS      TRFTRPREDB 1823 bp DNA INV 26-APR-1993
DEFINITION Crithidia fasciculata trypanothione reductase (tpr) gene (clone
TR5), complete cds.
ACCESSION M73324
VERSION M73324.1 GI:162504
KEYWORDS FAD oxidoreductase; NADPH cofactor; cytosolic enzyme; trypanothione
reductase.
SOURCE      Crithidia fasciculata DNA.
ORGANISM    Crithidia fasciculata
REFERENCE   1 (bases 1 to 1823)
AUTHORS    Field,H., Cerami,A. and Henderson,G.B.
TITLE       Cloning, sequencing and demonstration of polymorphism in
            trypanothione reductase from Crithidia fasciculata
JOURNAL     Mol. Biochem. Parasitol. 50, 47-56 (1992)
MEDLINE     92178290
FEATURES
  source     Location/Qualifiers
            1..1823
            /organism="Crithidia fasciculata"
            /db_xref="taxon:5656"
            /gene="tpr"
            /codon_start=1
            /product="trypanothione reductase"
            /protein_id="AAA30323.1"
            /db_xref="GI:162507"
            /translation="MSRAYDLVVIGAGSRLEAGWNAASLHKRVAVIDLQKHGPPH
            YALGGTCVNVGCVPRKKLVMTGANYMTDITRESAGFGWELDRSVRPNWKALIAARKNA
            VSGINDSYEGMFADTEGLTFHQFGALQDNHTLVRESADPNASVLETDTEYLILAT
            GSWPQHLGIEGDDLCITISNEAFYLDEAPKRALCVGGYISIEFAGIFNAYKARGGOVD
            LAYRGMILRGFDSRLKQLTEQLRANGINVRTHENPAKVTKNADGTRHVPESGAE
            DVDVYMLAIGRVPRSQTLQLEKAGVEYAKNGAIKYDAVSKTVNDNIYAIAGDVTDRYML
            TPVAINEGAAFDVTFANKPRATDHKKVACAFVSIPPMGVCYIYEEDAARKKIDQVAVY
            ESSFTPLMHNISGSTYKKFMVRITVNHADGEVLGVHMLGDSPEIIQISVAICLRMGAK
            ISDVNTTNGVHTPSAEELCSMTPTPAYFYQKGRVKEIDSNL"
            /translation="MSRAYDLVVIGAGSRLEAGWNAASLHKRVAVIDLQKHGPPH
            YALGGTCVNVGCVPRKKLVMTGANYMTDITRESAGFGWELDRSVRPNWKALIAARKNA
            VSGINDSYEGMFADTEGLTFHQFGALQDNHTLVRESADPNASVLETDTEYLILAT
            GSWPQHLGIEGDDLCITISNEAFYLDEAPKRALCVGGYISIEFAGIFNAYKARGGOVD
            LAYRGMILRGFDSRLKQLTEQLRANGINVRTHENPAKVTKNADGTRHVPESGAE
            DVDVYMLAIGRVPRSQTLQLEKAGVEYAKNGAIKYDAVSKTVNDNIYAIAGDVTDRYML
            TPVAINEGAAFDVTFANKPRATDHKKVACAFVSIPPMGVCYIYEEDAARKKIDQVAVY
            ESSFTPLMHNISGSTYKKFMVRITVNHADGEVLGVHMLGDSPEIIQISVAICLRMGAK
            ISDVNTTNGVHTPSAEELCSMTPTPAYFYQKGRVKEIDSNL"

BASE COUNT       372 a 596 c 559 g 296 t
ORIGIN

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-332-522B-2 x TRFTRPREDB ..
  Align seg 1/1 to: TRFTRPREDB from: 1 to: 1823

745 LeuArgGalaArgArgHisValArg 753
|||||
251 CTACGCCGCGCTCGCGCACGTCGCT 277

seq_name: gb_in1:TRFTRPREDC

seq_documentation_block:
LOCUS      TRFTRPREDC 1823 bp DNA INV 26-APR-1993
DEFINITION Crithidia fasciculata trypanothione reductase (tpr) gene (clone
TR1), complete cds.
ACCESSION M73325
VERSION M73325.1 GI:162506
KEYWORDS FAD oxidoreductase; NADPH cofactor; cytosolic enzyme; trypanothione
reductase.
SOURCE      Crithidia fasciculata DNA.
ORGANISM    Crithidia fasciculata
REFERENCE   1 (bases 1 to 1823)
AUTHORS    Field,H., Cerami,A. and Henderson,G.B.
TITLE       Cloning, sequencing and demonstration of polymorphism in
            trypanothione reductase from Crithidia fasciculata
JOURNAL     Mol. Biochem. Parasitol. 50, 47-56 (1992)
MEDLINE     92178290
FEATURES
  source     Location/Qualifiers
            1..1823
            /organism="Crithidia fasciculata"
            /db_xref="taxon:5656"
            /gene="tpr"
            /codon_start=1
            /product="trypanothione reductase"
            /protein_id="AAA30323.1"
            /db_xref="GI:162507"
            /translation="MSRAYDLVVIGAGSRLEAGWNAASLHKRVAVIDLQKHGPPH
            YALGGTCVNVGCVPRKKLVMTGANYMTDITRESAGFGWELDRSVRPNWKALIAARKNA
            VSGINDSYEGMFADTEGLTFHQFGALQDNHTLVRESADPNASVLETDTEYLILAT
            GSWPQHLGIEGDDLCITISNEAFYLDEAPKRALCVGGYISIEFAGIFNAYKARGGOVD
            LAYRGMILRGFDSRLKQLTEQLRANGINVRTHENPAKVTKNADGTRHVPESGAE
            DVDVYMLAIGRVPRSQTLQLEKAGVEYAKNGAIKYDAVSKTVNDNIYAIAGDVTDRYML
            TPVAINEGAAFDVTFANKPRATDHKKVACAFVSIPPMGVCYIYEEDAARKKIDQVAVY
            ESSFTPLMHNISGSTYKKFMVRITVNHADGEVLGVHMLGDSPEIIQISVAICLRMGAK
            ISDVNTTNGVHTPSAEELCSMTPTPAYFYQKGRVKEIDSNL"
            /translation="MSRAYDLVVIGAGSRLEAGWNAASLHKRVAVIDLQKHGPPH
            YALGGTCVNVGCVPRKKLVMTGANYMTDITRESAGFGWELDRSVRPNWKALIAARKNA
            VSGINDSYEGMFADTEGLTFHQFGALQDNHTLVRESADPNASVLETDTEYLILAT
            GSWPQHLGIEGDDLCITISNEAFYLDEAPKRALCVGGYISIEFAGIFNAYKARGGOVD
            LAYRGMILRGFDSRLKQLTEQLRANGINVRTHENPAKVTKNADGTRHVPESGAE
            DVDVYMLAIGRVPRSQTLQLEKAGVEYAKNGAIKYDAVSKTVNDNIYAIAGDVTDRYML
            TPVAINEGAAFDVTFANKPRATDHKKVACAFVSIPPMGVCYIYEEDAARKKIDQVAVY
            ESSFTPLMHNISGSTYKKFMVRITVNHADGEVLGVHMLGDSPEIIQISVAICLRMGAK
            ISDVNTTNGVHTPSAEELCSMTPTPAYFYQKGRVKEIDSNL"

```

seq_documentation_block: 1SDVNTICVHPTSAEELCSMRTPAFYEKGRVEKIDSNL"
LOCUS 369 a 592 c 564 g 298 t
DEFINITION Mouse immunoglobulin J558 V(H) gene VAR104 5'-flanking region.
ACCESSION X06856 M17571
VERSION X06856.1 GI:52440
KEYWORDS Ig heavy chain; immunoglobulin; pseudogene.
SOURCE house mouse.
ORGANISM Mus musculus

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x TRFTRPREDC ..

Align seg 1/1 to: TRFTRPREDC from: 1 to: 1823

745 LeuArgGArgAlaArgGHISValArg 753

|||||

251 CTACGGCGCGCTCGGGCACGTGCGT 277

seq_name: gb_ov:AF060951

seq_documentation_block:
LOCUS AF060951 2068 bp DNA VRT 12-MAY-1998
DEFINITION Gallus gallus homeobox C9 protein (Hoxc9) gene, intron 1.
ACCESSION AF060951
VERSION AF060951.1 GI:3127115
KEYWORDS Chicken.
SOURCE Gallus gallus
ORGANISM Gallus gallus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
AUTHORS 1 (bases 1 to 2068)
Papenbrock,T., Peterson,R.L., Lee,R.S., Hsu,T., Kuroiwa,A. and
Awgulewitsch,A.
TITLE The murine Hoxc9 gene contains a structurally and functionally
conserved enhancer
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2068)
Papenbrock,T., Peterson,R.L., Lee,R.S., Hsu,T., Kuroiwa,A. and
Awgulewitsch,A.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1998) Medicine/Rheumatology, Medical University
of South Carolina, 171 Ashley Ave., Charleston, SC 29425, USA
FEATURES
source
1..2068
/organism="Gallus gallus"
/db_xref="taxon:9031"
<1..>2068
/gene="Hoxc9"
/note="Hoxc9"
1..2068
/gene="Hoxc9"
/number=1

BASE COUNT 511 a 419 c 686 g 451 t 1 others
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x AF060951/rev ..

Align seg 1/1 to reverse of: AF060951 from: 1 to: 2068

227 ProProHishishishishis 235

|||||

1121 CCCCTCCTCACCACCACCACCACCAC 1095

seq_name: gb_ro:MMIGVH23

seq_documentation_block:
LOCUS MMIGVH23 2115 bp DNA ROD 07-MAY-1992
DEFINITION Mouse immunoglobulin J558 V(H) gene VAR104 5'-flanking region.
ACCESSION X06856 M17571
VERSION X06856.1 GI:52440
KEYWORDS Ig heavy chain; immunoglobulin; pseudogene.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Blankenstein,T.
TITLE Direct Submission
JOURNAL Submitted (11-FEB-1988) Blankenstein T., Institut fuer Genetik der
Universitaet zu Koeln, Weyertal 121, D-5000 Koeln 41, FRG
REFERENCE 2 (bases 1 to 2115)
Blankenstein,T., Bonhomme,F. and Krawinkel,U.
AUTHORS Evolution of pseudogenes in the immunoglobulin VH-gene family of
the mouse
JOURNAL Immunogenetics 26 (4-5), 237-248 (1987)
MEDLINE 88006305
COMMENT VAR104 is present in Balb/c genome (southern blot).
FEATURES
source
1..2115
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="B1-8.V1"
/clone_lib="Charon 30."
/clone="V104"

misc_feature 1..2115
/note="VAR104 5'-flanking region"
BASE COUNT 673 a 359 c 379 g 704 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x MMIGVH23 ..

Align seg 1/1 to: MMIGVH23 from: 1 to: 2115

227 ProProHishishishishis 235
|||||

1898 CCTCCTCCTCATCATCATCATCAT 1924

seq_name: gb_ro:MUSIGHYB1

seq_documentation_block:
LOCUS MUSIGHYB1 2117 bp DNA ROD 27-APR-1993
DEFINITION Mouse Ig heavy-chain variable region (V) pseudogene VAR104 DNA, 5'
flanking region.
ACCESSION M17571
VERSION M17571.1 GI:196237
KEYWORDS V-region; immunoglobulin heavy chain; pseudogene.
SEGMENT 1 of 4
SOURCE Mouse DNA from hybridoma B1-8 library.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 2117)
Blankenstein,T., Bonhomme,F. and Krawinkel,U.
TITLE Evolution of pseudogenes in the immunoglobulin VH-gene family of
the mouse
JOURNAL Immunogenetics 26, 237-248 (1987)
MEDLINE 88006305
FEATURES
source
1..2117
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"

Thu Sep 28 07:39:48 2000

```
BASE COUNT 674 a 359 c 379 g 705 t
ORIGIN BamHI site.

alignment_scores:
  Quality: 9.00 Length: 9
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x MUSIGHYB1 ..
Align seg 1/1 to: MUSIGHYB1 from: 1 to: 2117

227 ProProHisHisHisHisHisHis 235
1900 CCTCCTCCTCATCATCATCATCAT 1926
```

THIS PAGE BLANK (USPTO)

OM of: US-09-332-522B-2_COPY_335_428 to: N_Geneseq_36:* out_format : pfs

Date: Sep 27, 2000 2:46 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+pn2.model -DEV=xlp
-Q/cgn2.1/USPTO_spool/US09332522/runat_27092000_091222_6568/app_query.fasta_1.1355
-DB=N_Geneseq_36 -QPMT=fastap -SUFFIX=oli.rng -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -CGAPEXT=0.050 -GAPOP=60.000 -XGAPEXT=60.000
-FCGAPOP=6.000 -FCGAPEXT=7.000 -GAPOP=60.000 -XGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09332522_@CGN1_1_168 -NCPU=6 -ICPU=3
-LOGLOG -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-332-522B-2_COPY_335_428

Query length: 94

Database: N_Geneseq_36:*

Database sequences: 311585

Database length: 125096042

Search time (sec): 112.180000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000

WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

Sequence	Strid	Orig	ZScore	Escore	Len	Documentation
N_Geneseq_36:V89537	+	8.00	147.72	1.26	122	EST clone CQ309. New polynucleotide acid encoding interleukin-2
N_Geneseq_36:V13185	+	7.00	112.45	115.76	1188	Nucleic acid encoding interleukin-2
N_Geneseq_36:V13185	+	7.00	110.84	142.31	1481	Nucleic acid encoding interleukin-2
N_Geneseq_36:V13185	+	7.00	108.37	195.44	2078	Human secreted protein cDNA fragment
N_Geneseq_36:V13185	+	7.00	107.80	195.96	2084	Human secreted protein cDNA fragment
N_Geneseq_36:V13185	+	7.00	107.80	210.11	2245	Human secreted protein cDNA fragment
N_Geneseq_36:V13185	+	7.00	106.95	234.17	2522	Clone df396.1 isolated from human
N_Geneseq_36:V13185	+	7.00	106.60	245.17	2647	DNA - b sequence of Bean Golden
N_Geneseq_36:V13185	+	7.00	106.60	245.17	2647	Vector sequence derived from Bean
N_Geneseq_36:V13185	+	7.00	105.36	287.19	3134	Leishmania major M15 antigen cDNA
N_Geneseq_36:V13185	+	7.00	105.36	287.19	3134	Leishmania antigen M15 coding
N_Geneseq_36:V13185	+	7.00	92.66	1.5e+03	17846	Staphylococcus aureus contig
N_Geneseq_36:V13185	+	7.00	81.70	6.0e+03	80073	Continuation (6 of 6) of T58
N_Geneseq_36:V13185	+	7.00	79.38	8.0e+03	110000	Continuation (12 of 17) of T58
N_Geneseq_36:V13185	+	6.00	123.63	27.59	20	Maize oligonucleotide marker S44
N_Geneseq_36:V13185	+	6.00	123.28	28.88	21	Primer used in detection of HPV
N_Geneseq_36:V13185	+	6.00	122.30	32.72	24	Primer PRC02-25 mer (sic) for th
N_Geneseq_36:V13185	+	6.00	122.30	32.72	24	Primer 2 to amplify 300bp bacter
N_Geneseq_36:V13185	+	6.00	122.00	34.00	25	PCR primer 2 for phage lambda DN
N_Geneseq_36:V13185	+	6.00	122.00	34.00	25	PCR primer PCR02. Homogeneous an
N_Geneseq_36:V13185	+	6.00	122.00	34.00	25	PCR primer used to demonstrate a
N_Geneseq_36:V13185	+	6.00	122.00	34.00	25	PCR antisense primer for the amp
N_Geneseq_36:V13185	+	6.00	122.00	34.00	25	5'-3' sequence of inverse linkag
N_Geneseq_36:V13185	+	6.00	122.00	34.00	25	3'-5' sequence of inverse linkag
N_Geneseq_36:V13185	+	6.00	122.00	34.00	25	M13mp18 phage DNA primer. Assay
N_Geneseq_36:V13185	+	6.00	122.00	34.00	25	Primer PCR02 for bacteriophage l
N_Geneseq_36:V13185	+	6.00	122.00	34.00	25	Bacteriophage lambda specific se
N_Geneseq_36:V13185	+	6.00	122.00	34.00	25	Bacteriophage lambda genome prim
N_Geneseq_36:V13185	+	6.00	122.00	34.00	25	Oligonucleotide of the invention
N_Geneseq_36:V13185	+	6.00	122.00	34.00	25	Primer #2 for bacteriophage lamb
N_Geneseq_36:V13185	+	6.00	122.00	34.00	25	PCR primer SEQ ID NO:3 from US58
N_Geneseq_36:V13185	+	6.00	121.72	35.27	26	PCR primer SEQ ID NO:1 from US58
N_Geneseq_36:V13185	+	6.00	120.67	40.33	30	Bacillus amyloliquefaciens subti
N_Geneseq_36:V13185	+	6.00	120.67	40.33	30	Primer used to amplify the subti
N_Geneseq_36:V13185	+	6.00	118.94	50.33	38	Nucleotide sequence of sequence
N_Geneseq_36:V13185	+	6.00	118.57	52.80	40	Plasmid p182Sfl component oligo
N_Geneseq_36:V13185	+	6.00	118.57	52.80	40	3' PCR primer for linking human
N_Geneseq_36:V13185	+	6.00	118.39	54.04	41	Primer PRC02A-40 mer for the lam

N_Geneseq_36:V89537	-	6.00	117.55	60.19	46	Differentially expressed pre-
N_Geneseq_36:V15666	-	6.00 <td>117.55<td>60.19<td>46<th>PCR primer sequence. Multiple</th></td></td></td>	117.55 <td>60.19<td>46<th>PCR primer sequence. Multiple</th></td></td>	60.19 <td>46<th>PCR primer sequence. Multiple</th></td>	46 <th>PCR primer sequence. Multiple</th>	PCR primer sequence. Multiple
N_Geneseq_36:V20001	-	6.00 <td>117.24<td>62.64<td>48<th>PCR primer for mutated KGF-2</th></td></td></td>	117.24 <td>62.64<td>48<th>PCR primer for mutated KGF-2</th></td></td>	62.64 <td>48<th>PCR primer for mutated KGF-2</th></td>	48 <th>PCR primer for mutated KGF-2</th>	PCR primer for mutated KGF-2
N_Geneseq_36:Q23242	+	6.00 <td>117.09<td>63.86<td>49<th>Lambda bacteriophage PCR prim</th></td></td></td>	117.09 <td>63.86<td>49<th>Lambda bacteriophage PCR prim</th></td></td>	63.86 <td>49<th>Lambda bacteriophage PCR prim</th></td>	49 <th>Lambda bacteriophage PCR prim</th>	Lambda bacteriophage PCR prim
N_Geneseq_36:N91957	-	6.00 <td>116.94<td>65.08<td>50<th>Complementary strand of the N</th></td></td></td>	116.94 <td>65.08<td>50<th>Complementary strand of the N</th></td></td>	65.08 <td>50<th>Complementary strand of the N</th></td>	50 <th>Complementary strand of the N</th>	Complementary strand of the N
N_Geneseq_36:Q61446	+	6.00 <td>116.94<td>65.08<td>50<th>Bacteriophage lambda ligand D</th></td></td></td>	116.94 <td>65.08<td>50<th>Bacteriophage lambda ligand D</th></td></td>	65.08 <td>50<th>Bacteriophage lambda ligand D</th></td>	50 <th>Bacteriophage lambda ligand D</th>	Bacteriophage lambda ligand D

seq_name: N_Geneseq_36:V89537

seq_documentation_block:

ID V89537 standard; cDNA; 122 BP.
AC V89537;
DT 15-FEB-1999 (first entry)
DE EST clone CQ309.
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW gene therapy; ss.
OS Homo sapiens.
PN WO9845436-A2.
PD 15-OCT-1998.
PF 10-APR-1998; U06955.
PR 10-APR-1997; US-838821.
PA (GENY) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 99-070077/06.
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
PS Claim 1; Page 249; 618pp; English.
CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
SQ Sequence 122 BP; 40 A; 22 C; 24 G; 36 T;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x V89537 ..
Align seg 1/1 to: V89537 from: 1 to: 122

41 ileGlnGlnLeuLysValLeuLeu 48

|||||

96 ATACACCAATGGAAGTGTACTC 119

seq_name: N_Geneseq_36:T31855

seq_documentation_block:

ID T31855 standard; DNA; 1188 BP.
AC T31855;
DT 16-JAN-1997 (first entry)
DE Nucleic acid encoding interleukin-2 receptor associated protein p43.
KW p43; interleukin-2 receptor; IL-2; treatment; diagnosis; disease;
KW signal transduction; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 1..1188
FT /*tag= a

```

FT      WO9621732-A1.
PN      /product= p43 polypeptide.
PD      18-JUL-1996.
PF      23-DEC-1995; E05123.
PR      09-JAN-1995; EP-100201.
PA      (BOEH ) BOEHRINGER INGELHEIM INT GMBH.
PI      Barsomian EL, Shibuya H, Taniguchi T;
DR      WPI: 96-342285/34.
DR      P-PSDB; R97569.
PT      p43 and related polypeptide(s) which bind to interleukin receptor
PT      - useful for treatment and diagnosis of IL-2 related disease, and to
PT      study mechanism of IL-2 mediated signal transduction
PS      Claim 11; Page 25-26; 39pp; English.
CC      A polypeptide having p43-like biological activity, related peptides
CC      and antisense oligonucleotides to the p43 gene are useful in the
CC      treatment and diagnosis of interleukin-2 related diseases. p43 can
CC      also be used to study the mechanism of IL-2 mediated signal
CC      transduction.
SQ      Sequence 1188 BP; 333 A; 226 C; 298 G; 331 T;

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x T31855
Align seg 1/1 to: T31855 from: 1 to: 1188

52 GluAlaLysLeuSerIysSer 58
|||||
62 GAAGCAAAAGTTGCCAAAGC 82

seq_name: N_Geneseq_36:T31854

seq_documentation_block:
ID T31854 standard; DNA; 1481 BP.
AC T31854;
DE Nucleic acid encoding interleukin-2 receptor associated protein p43.
KW p43; interleukin-2 receptor; IL-2; treatment; diagnosis; disease;
KW signal transduction; ss.
OS Homo sapiens.
FH Key
FT Location/Qualifiers
FT 246..1436
FT /*tag= a
FT /*product= p43 polypeptide.

PN WO9621732-A1.
PD      18-JUL-1996.
PF      23-DEC-1995; E05123.
PR      09-JAN-1995; EP-100201.
PA      (BOEH ) BOEHRINGER INGELHEIM INT GMBH.
PI      Barsomian EL, Shibuya H, Taniguchi T;
DR      WPI: 96-342285/34.
DR      P-PSDB; R97569.
PT      p43 and related polypeptide(s) which bind to interleukin receptor
PT      - useful for treatment and diagnosis of IL-2 related disease, and to
PT      study mechanism of IL-2 mediated signal transduction
PS      Claim 11; Page 24-25; 39pp; English.
CC      A polypeptide having p43-like biological activity, related peptides
CC      and antisense oligonucleotides to the p43 gene are useful in the
CC      treatment and diagnosis of interleukin-2 related diseases. p43 can
CC      also be used to study the mechanism of IL-2 mediated signal
CC      transduction.
SQ      Sequence 1481 BP; 432 A; 271 C; 357 G; 421 T;

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

```

```

alignment_block:
US-09-332-522B-2_COPY_335_428 x T31854
Align seg 1/1 to: T31854 from: 1 to: 1481

52 GluAlaLysLeuSerIysSer 58
|||||
307 GAAGCAAAAGTTGCCAAAGC 327

seq_name: N_Geneseq_36:X37439

seq_documentation_block:
ID X37439 standard; cDNA; 2078 BP.
AC X37439;
DE 06-JUL-1999 (first entry)
DE Human secreted protein cDNA fragment containing gene 11.
KW Human; secreted protein; prevention; treatment; protein therapy;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; foetal deficiency; blood disorder; lymphoma;
KW leukemia; immune system disorder; autoimmune disease; hepatic disease;
KW renal disease; inflammation; allergy; asthma; sepsis; diabetes; AIDS;
KW Alzheimer's disease; cognitive disorder; schizophrenia; osteoporosis;
KW arthritis; psoriasis; digestive; endocrine; infection; ss.
OS Homo sapiens.
PN WO9909155-A1.
PD 25-FEB-1999.
PF 18-AUG-1998; US-092956.
PR 16-JUN-1998; US-092956.
PR 15-JUL-1998; US-092956.
PR 19-AUG-1997; US-056368.
PR 19-AUG-1997; US-056369.
PR 19-AUG-1997; US-056335.
PR 19-AUG-1997; US-056555.
PR 19-AUG-1997; US-056556.
PR 19-AUG-1997; US-056628.
PR 19-AUG-1997; US-056629.
PR 19-AUG-1997; US-056726.
PR 19-AUG-1997; US-056728.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Duan R, Ebner R, Endress GA, Feng P,
PI Florence C, Florence KA, Komatsoulis GA, Lafleur DW,
PI Moore PA, Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR,
PI Young PE;
DR WPI: 99-190160/16.
DR P-PSDB; Y07814.
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1a; Page 230; 280pp; English.
CC This invention describes novel isolated human secreted proteins and
CC their encoding nucleic acid sequences. The products of the invention
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also pathological conditions can be
CC diagnosed by determining the presence or amount of expression of
CC the new polypeptides in a sample or by determining the presence or
CC absence of mutations in the new polynucleotides. Specific uses are
CC described for each of the 70 polynucleotides, based on which tissues they
CC are most highly expressed in, and include developing products for the
CC diagnosis or treatment of cancer, tumours, neurodegenerative
CC disorders, developmental abnormalities and foetal deficiencies, blood
CC disorders, leukemias, diseases of the immune system, autoimmune diseases,
CC hepatic and renal disease, lymphomas, inflammation, allergies, asthma,
CC sepsis, diabetes, Alzheimer's and cognitive disorders, schizophrenia,
CC osteoporosis, arthritis, psoriasis, digestive/endocrine disorders,
CC infections and AIDS. The human secreted proteins of the invention are
CC represented in Y07744-Y07850 and the encoding nucleic acids are
SQ Sequence 2078 BP; 429 A; 566 C; 579 G; 497 T;

alignment_scores:
  Quality: 7.00      Length: 7

```

Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x X37439/rev ..

Align seg 1/1 to reverse of: X37439 from: 1 to: 2078

41 IleglnGlnLeuLysValleu 47
|||||
1032 ATCCAGCACTCAAGTCCTC 1012

seq_name: N_Geneseq_36:X37379

seq_documentation_block:

ID X37379 standard; cDNA; 2084 BP.
AC X37379;
DT 06-JUL-1999 (first entry)
DE Human secreted protein cDNA fragment containing gene 11.
KW Human; secreted protein; cDNA; prevention; treatment; protein therapy;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; foetal deficiency; blood disorder; lymphoma;
KW leukemia; immune system disorder; autoimmune disease; hepatic disease;
KW renal disease; inflammation; allergy; asthma; sepsis; diabetes; AIDS;
KW Alzheimer's disease; cognitive disorder; schizophrenia; osteoporosis;
KW arthritis; psoriasis; digestive; endocrine; infection; ss.
OS Homo sapiens.
PN WO909155-A1.
PD 25-FEB-1999.
PF 18-AUG-1998: U17044.
PR 16-JUN-1998: US-092956.
PR 15-JUL-1998: US-092956.
PR 19-AUG-1997: US-056368.
PR 19-AUG-1997: US-056369.
PR 19-AUG-1997: US-056535.
PR 19-AUG-1997: US-056555.
PR 19-AUG-1997: US-056556.
PR 19-AUG-1997: US-056628.
PR 19-AUG-1997: US-056629.
PR 19-AUG-1997: US-056726.
PR 19-AUG-1997: US-056728.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Duan R, Ebner R, Endress GA, Feng P,
PI Florence C, Florence KA, Komatsoulis GA, Lafleur DW,
PI Moore PA, Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR,
PI Young PE;
DR WPI: 99-190160/16.
DR P-PSDB: Y07754.
PT New isolated human genes and the secreted polypeptides they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1a: Page 196; 280pp: English.
CC This invention describes novel isolated human secreted proteins and
CC their encoding nucleic acid sequences. The products of the invention
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also pathological conditions can be
CC diagnosed by determining the presence or amount of expression of
CC the new polypeptides in a sample or by determining the presence or
CC absence of mutations in the new polynucleotides. Specific uses are
CC described for each of the 70 polynucleotides, based on which tissues they
CC are most highly expressed in, and include developing products for the
CC diagnosis or treatment of cancer, tumours, neurodegenerative
CC disorders, developmental abnormalities and foetal deficiencies, blood
CC disorders, leukemias, diseases of the immune system, autoimmune diseases,
CC hepatic and renal disease, lymphomas, inflammation, allergies, asthma,
CC sepsis, diabetes, Alzheimer's and cognitive disorders, schizophrenia,
CC osteoporosis, arthritis, psoriasis, digestive/endocrine disorders,
CC infections and AIDS. The human secreted proteins of the invention are
CC represented in Y07744-Y07850 and the encoding nucleic acids are
CC represented in X37369-X37441.
SQ Sequence 2084 BP; 436 A; 569 C; 584 G; 492 T;

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x X37379/rev ..

Align seg 1/1 to reverse of: X37379 from: 1 to: 2084

41 IleglnGlnLeuLysValleu 47
|||||
1031 ATCCAGCACTCAAGTCCTC 1011

seq_name: N_Geneseq_36:V61485

seq_documentation_block:

ID V61485 standard; cDNA; 2245 BP.
AC V61485;
DT 11-JAN-1999 (first entry)
DE Human secreted protein do568_11 cDNA.
KW Secreted protein; human; do568_11; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 359..1369
FT /tag= a
PN WO9841539-A2.
PD 24-SEP-1998.
PF 19-MAR-1998: U05474.
PR 18-MAR-1998: US-040963.
PR 19-MAR-1997: US-820493.
PA (GEMI) GENETICS INST INC.
PI Agostino MJ, Jacobs K, LaVallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI: 98-521163/44.
DR P-PSDB: W79094.
PT New polynucleotide(s) encoding secreted human proteins - derived
PT from human foetal kidney, adult testes and adult or foetal brain
PT cDNA libraries
PS Claim 30; Page 82-84; 112pp: English.
CC This full-length cDNA clone, designated do568_11, codes for a novel
CC secreted human protein (see W79094). It was isolated from a human
CC adult testis cDNA library using methods which are selective for
CC cDNAs encoding secreted proteins, or was identified as encoding
CC a secreted or transmembrane protein on the basis of computer
CC analysis of the encoding protein. The nucleotide sequence shows
CC homology to some database sequences. The invention provides
CC cDNA clones (see V61477-87) from human foetal kidney, adult testis,
CC and adult or foetal brain cDNA libraries that code for secreted
CC proteins (see W79087-97). These clones are deposited as
CC ATCC 98364. The polynucleotides and proteins are predicted to have
CC useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans
CC and animals, although no supporting data is given. Suggested
CC activities include nutritional, immune stimulating (e.g. as
CC vaccines) or suppressing, haematopoiesis regulating, tissue growth,
CC activin/inhibin, chemotactic/chemokinetic, haemostatic and
CC thrombolytic, receptor/ligand, antiinflammatory, cadherin/tumour
CC invasion suppressor and tumour inhibition activities. The
CC polynucleotides are also stated to be useful for gene therapy.
SQ Sequence 2245 BP; 443 A; 632 C; 617 G; 546 T;

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x V61485/rev ..

Align seg 1/1 to reverse of: V61485 from: 1 to: 2245

41 lleGInLeuLysValLeu 47
 1225 ATCCAGCACTCAAGTCTC 1205

seq_name: N_Geneseq_36:V82781

seq_documentation_block:
 ID V82781 standard; CDNA; 2522 BP.

AC V82781;
 DE 25-FEB-1999 (first entry)
 DE Clone df396.1 isolated from human adult brain cDNA library.
 DE Secreted protein: nutritional activity; immune stimulating; vaccine;
 KW suppressing activity; haematopoiesis regulating activity;
 KW tissue growth activity; activin; inhibin activity; chemotaxis;
 KW chemokinetic activity; haemostasis; thrombolytic activity; receptor;
 KW ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
 KW tumour inhibition; gene therapy; ds.
 OS Homo sapiens.
 PN WQ9842739-A2.
 PN 01-OCT-1998.
 PF 20-MAR-1998; U05653.
 PR 19-MAR-1998; US-044466.
 PR 21-MAR-1997; US-822167.
 PA (GEMV) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
 PI Racie LA, Spaulding V, Treacy M;
 DR WPI: 98-609890/51.
 DR P-PSDB; W85458.
 DR New polynucleotides encoding secreted human proteins - derived from
 PT human foetal brain, adult brain, foetal kidney, placenta or adult
 PT pineal gland cDNA libraries.
 PS Claim 20; Page 74-76; 11pp; English.
 CC The present sequence encodes a secreted protein. The polynucleotide and
 CC secreted protein are predicted to have biological activities which would
 CC make them suitable for treating, preventing or ameliorating medical
 CC conditions in humans and animals, although no supporting data is given.
 CC Suggested activities include nutritional activity, immune stimulating
 CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
 CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 CC invasion suppressor activity, and tumour inhibition activity (no data is
 CC given in the specification to support these activities). The
 CC polynucleotide is also stated to be useful for gene therapy.
 SQ Sequence 2522 BP; 558 A; 683 C; 608 G; 673 T;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x V82781/rev ..

Align seg 1/1 to reverse of: V82781 from: 1 to: 2522

57 LysSerAlaThrLeuArgArg 63
 1033 AAAAGTGCACTTGAGAAGA 1013

seq_name: N_Geneseq_36:N70630

seq_documentation_block:
 ID N70630 standard; DNA; 2647 BP.

AC N70630;
 DE 02-MAY-1991 (first entry)
 DE DNA -b sequence of Bean Golden Mosaic virus.
 KW Plant vector; ds.
 OS Bean golden mosaic virus.
 FH Key Location/Qualifiers
 FT cds complement (4. .1595)
 FT /*tag= a

FT /label= ORF 4
 FT /note= "Connects with 3'-terminal"
 FT 340. .1125
 FT /*tag= b
 FT /label= ORF 3
 FT 344. .688
 FT /*tag= c
 FT /label= ORF 6
 FT complement (703. .1020)
 FT /*tag= d
 FT /label= ORF 7
 FT complement (1122. .1520)
 FT /*tag= e
 FT /label= ORF 8
 FT complement (1267. .1785)
 FT /*tag= f
 FT /label= ORF 5
 FT complement (2239. .2496)
 FT /*tag= g
 FT /label= ORF 9
 FT J61257186-A.
 FT 14-NOV-1986.
 PD 10-MAY-1985; 098108.
 PF 10-MAY-1985; JP-098108.
 PR (TEIJ) TEIJIN KK.
 PA WPI: 87-159662/23.
 DR P-PSDB; P70406-12.
 DR New DNA and hybrid DNA - used for recombinant vector of plants.
 PT Claim 4; Fig 2; 24pp; Japanese.
 PS The sequence may be taken along with the -a DNA sequence and a
 CC suitable resistance gene, and used to create a recombinant plant
 CC vector.
 CC See also N70629.
 CC Sequence 2647 BP; 707 A; 521 C; 594 G; 825 T;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x N70630/rev ..

Align seg 1/1 to reverse of: N70630 from: 1 to: 2647

78 LeuLysHisValGluGln 84
 225 TTAAGCACCCACGTGGAGCAG 205

seq_name: N_Geneseq_36:N60611

seq_documentation_block:
 ID N60611 standard; DNA; 2647 BP.

AC N60611;
 DT 16-OCT-1991 (first entry)
 DE Vector sequence derived from bean golden mosaic virus.
 KW BGW; haricot beans; ds.
 OS Bean golden mosaic virus.
 PN J61132188-A.
 PD 19-JUN-1986.
 PF 03-DEC-1984; 254185.
 PR 03-DEC-1984; JP-254185.
 PA (TEIJ) TEIJIN KK.
 DR WPI: 86-200544/31.
 PT Novel DNA and hybrid DNA - derived from bean golden mosaic virus.
 PS Claim 4; table 2; 17pp; Japanese.
 CC Vector is derived from the BGW genome, and is useful recombination
 CC of plant genes. Sequence is referred to in the patent as DNA-B. See
 CC also N60610.
 SQ Sequence 2647 BP; 706 A; 519 C; 597 G; 825 T;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x V21209_11/rev ..

Align seg 1/1 to reverse of: V21209_11 from: 1 to: 110000

60 ThrLeuArgGAlaIleGlu 66
|||||

39202 ACTCTAAGAGGAGCAATGAA 39182

seq_name: N_Geneseq_36:V40349

seq_documentation_block:
ID V40349 standard; DNA; 20 BP.
AC V40349;
DT 14-OCT-1998 (first entry)
DE Maize oligonucleotide marker S44R.
KW Maize; marker; probe; PCR primer; polymorphism; vegetal sequence;
KW polymorphic site; corn; gramineae species; ss.
OS Synthetic.
OS Zea sp.
PN W09830717-A2.
PD 16-JUL-1998.
PF 02-DEC-1997; E07134.
PR 02-DEC-1996; US-032069.
PA (BIOC-) BIOCEM SA.
PI Murligneux A;
DR WPI; 98-399160/34.
PT Vegetal sequences including single nucleotide polymorphism - useful,
PT e.g. to determine polymorphisms in plants, determine strain in plant
PT breeding and to correlate polymorphisms with phenotypic traits
PS Example 2; Page 9; 32pp; English.
CC The present invention describes a nucleic acid segment comprising at
CC least 10 contiguous nucleotides from a vegetal sequence including a
CC polymorphic site which is a single nucleotide polymorphism (SNP), or the
CC complement of the segment. Also described are: (1) an allele-specific
CC oligonucleotides hybridising to segment, or their complements and (2) a
CC method of analysing nucleic acids from a subject, by determining if a
CC base is occupying any one (or a set) of polymorphic sites in 261
CC sequences derived from six maize lines (see V47701 to V47961). The
CC segments are useful in fingerprint analysis in plants to determine which
CC polymorphisms are present, which strain a plant belongs to and to
CC distinguish between strains. The polymorphisms may correlate with
CC phenotypic traits (e.g. plant growth rate or crop yield), and the
CC segments are useful to determine the presence/absence of specific
CC polymorphisms correlating with the existence/absence of particular
CC traits. The segments are also useful in marker assisted back-cross
CC techniques to select plants with a higher percentage of recurrent parent
CC in a back-cross population. Segments incorporate SNPs which occur more
CC frequently than other polymorphism types and are therefore more likely to
CC be located close to genetic loci of interest; different forms of
CC characterised SNPs are also often easier to detect than other
CC polymorphism types. V40304 to V40369 are used in an example from the
CC present invention as markers and PCR primers.
SQ Sequence 20 BP; 3 A; 6 C; 6 G; 5 T;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x V40349/rev ..

Align seg 1/1 to reverse of: V40349 from: 1 to: 20

58 SerAlaThrLeuArgArg 63
|||||

19 TCCGCACACACTGCCAAGG 2

seq_name: N_Geneseq_36:T32306

seq_documentation_block:
ID T32306 standard; DNA; 21 BP.
AC T32306;

DT 08-NOV-1996 (first entry)
DE Primer used in detection of HPV specific target sequence.
DE Detection; amplification; primer; probe; target sequence;
KW screening; human papilloma virus; HPV; cystic fibrosis; ss.
OS Synthetic.
PN W09615271-A1.
PD 23-MAY-1996.
PF 15-NOV-1995; U14886.
PR 16-NOV-1994; US-344203.
PA (ABBO) ABBOTT LAB.
PI Carrino JJ;
DR WPI; 96-259864/26.

PT Multiplex ligation dependent amplification - used to detect target
PT nucleic acid sequence and specifically amplify multiple target
PT sequences using single primer pair.
PS Example 1; Page 24; 38pp; English.
CC A method for the detection and amplification of a target sequence
CC (TS), comprises (A) forming a reaction mixture under hybridising
CC conditions with a sample suspected of containing the TS in a single
CC stranded form and at least one split probe (SP) having a 5' end
CC complementary to a first segment of the TS and a 3' end
CC complementary to second segment of the TS, both segments being
CC sufficiently close to allow ligation of the SP components
CC when they are hybridised to the TS; and where the SP components also
CC comprise a second, non-complementary region located
CC downstream of the 5' end and upstream of the 3' end. (B) Ligation
CC the 5' end and 3' end together and then separating the ligated
CC product from the TS. (C) treating the reaction mixture under
CC hybridising conditions with an excess of amplification primers which
CC bind to primer binding sites in the non-complementary regions of the
CC ligated SP. (D) Extending the amplification primers and then
CC treating the reaction mixture under denaturation conditions to
CC separate the amplification product from the ligated SP probe
CC template. (E) Returning to step (C) and repeating the procedure.
CC The method is used to amplify and detect a target nucleic acid
CC sequence and to specifically amplify multiple target sequences using
CC a single pair of primers. NOTE: The 5' and 3' ends of the SP may be
CC on two distinct polynucleotides or on different ends of one
CC continuous polynucleotide. The amplification primers used in the
CC detection of HPV TS using this method are given in T32305 and T32306.
CC A probe complementary to the amplified product is given in T32307.
CC See also T32302-04 for sequences of SP reagents.
SQ Sequence 21 BP; 7 A; 5 C; 5 G; 4 T;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x T32306 ..

Align seg 1/1 to: T32306 from: 1 to: 21

55 LeuSerLysSerAlaThr 60
|||||

3 TTATCGAATCAGCCACA 20

seq_name: N_Geneseq_36:Q14786

seq_documentation_block:
ID Q14786 standard; DNA; 24 BP.
AC Q14786;

DT 24-FEB-1992 (first entry)
 DE Primer PRC02-25 mer (sic) for the lambda viral genome.
 KW Homologous recombination; RecA; amplification; probe; RecA-803; ss.
 OS Synthetic.
 PN W09117267-A.
 PD 14-NOV-1991.
 PF 17-APR-1991; U02626.
 PR 07-MAY-1990; US-520321.
 PA (STRI) SRI INTERNATIONAL.
 PI Zarling DA, Sena EP, Green CJ;
 DR WPI; 91-353786/48.
 PT Amplification of target DNA sequences - using primers, RecA
 protein, ATP-gamma-S, dNTPs and DNA polymerase.
 PS Example 7; Page 28; 52pp; English.
 CC The sequence was used to demonstrate RecA+ protein enhancement of
 CC DNA synthesis on native linear lambda DNA templates. The primer
 CC corresponds to nucleotides 7606-7603 of the lambda viral genome.
 CC (it is a 24mer, despite its name). It was coated in RecA+ protein
 CC using ATP-gamma-S. The coated primer was then added to 0.5 ug
 CC lambda viral genomic DNA SSB protein. After 3 mins at 37 deg.C,
 CC 1 U Klenow fragment was added. Results showed that the addition of
 CC RecA+ protein clearly enhanced bulk DNA synthesis.
 CC See also Q14781-88.
 SQ Sequence 24 BP; 7 A; 7 C; 6 G; 4 T;

alignment_scores:
 Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x Q14786 ..
 Align seg 1/1 to: Q14786 from: 1 to: 24

55 LeuSerLysSerAlaThr 60
 |||||

3 TTATCGAAATCAGCCACA 20

seq_name: N_Geneseq_36:Q61451

seq_documentation_block:
 ID Q61451 standard; DNA; 24 BP.

AC Q61451;

DT 28-NOV-1994 (first entry)

DE Primer 2 to amplify 300bp bacteriophage lambda target sequence.

KW Detection; bacteriophage lambda; differential hybridisation;

KW elution profile; separation column; PCR primer;

KW polymerase chain reaction amplification; ss.

OS Synthetic.

PN J06078799-A.

PD 22-MAR-1994.

PF 31-AUG-1992; 257148.

PR 31-AUG-1992; JP-257148.

PA (SUME) SUMITOMO ELECTRIC IND CO.

DR WPI; 94-131295/16.

PT Detecting specific DNA fragments - comprises immobilising ligand

PT on insol. carrier used as solid phase in elution column,

PT increasing the temp. and analysing elution profile

PS Example 1; Page 5; 6pp; Japanese.

CC A sequence (Q61446) from bacteriophage lambda was synthesised and

CC immobilised on silica gel via "aminolink" bound to its 5'-end. The

CC sequence is an example of a ligand to be bound to a target DNA

CC fragment. A DNA sample (made single-stranded by denaturation) is

CC contacted with the immobilised ligand which is packed into a column.

CC Only the target DNA is specifically bound and other DNA passes

CC through the column. The target DNA fragment is eluted by increasing

CC the column temperature. Primers 1-1, 1-2 and 1-3 (Q61448-Q61450) were

CC each used with primer 2 (Q61451) to amplify a 300bp sequence from

CC lambda DNA. Primers 1-2 and 1-3 each introduced single base

CC mismatches into the target. The amplified fragments were used to

CC demonstrate the specificity of the novel detection method.

SQ Sequence 24 BP; 7 A; 7 C; 6 G; 4 T;

alignment_scores:

Quality: 6.00

Ratio: 1.000

Length: 6

Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x Q61451 ..

Align seg 1/1 to: Q61451 from: 1 to: 24

55 LeuSerLysSerAlaThr 60

|||||

3 TTATCGAAATCAGCCACA 20

seq_name: N_Geneseq_36:Q11620

seq_documentation_block:

ID Q11620 standard; DNA; 25 BP.

AC Q11620;

DT 03-JUL-1991 (first entry)

DE PCR primer 2 for phage lambda DNA.

KW Biopolymer; synthon; label; ss.

OS Synthetic.

PN EP-424819-A.

PD 02-MAY-1991.

PF 19-OCT-1990; 120093.

PR 23-OCT-1991; US-425740.

PA (MIFI) Millipore Corporation.

PI Coull JM, Galdea B, Koester H;

DR WPI; 91-126363/18.

PT Protecting natural prods, biopolymers or their synthons - used esp.

PT in prepn. of oligonucleotides.

PS Disclosure; Page 23; 32pp; English.

CC The primer can be reversibly modified by coupling it to a detection

CC label esp. fluorescein. See also Q11619.

SQ Sequence 25 BP; 7 A; 8 C; 6 G; 4 T;

alignment_scores:

Quality: 6.00

Ratio: 1.000

Length: 6

Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x Q11620 ..

Align seg 1/1 to: Q11620 from: 1 to: 25

55 LeuSerLysSerAlaThr 60

|||||

3 TTATCGAAATCAGCCACA 20

seq_name: N_Geneseq_36:Q21949

seq_documentation_block:

ID Q21949 standard; DNA; 25 BP.

AC Q21949;

DT 18-JUN-1992 (first entry)

DE PCR primer PCR02.

KW Amplification assay; probe; specificity; ss.

OS Synthetic.

PN W09202638-A.

PD 20-FEB-1992.

PF 06-AUG-1991; U05571.

PR 06-AUG-1990; US-563758.

PA (CETU) CETUS CORP.

PI Gelfand DH, Holland PM, Saiki RK, Watson RM;

DR WPI; 92-080091/10.

PT Homogeneous assay system for detecting target nucleic acid -

PT using labelled oligo:nucleotide(s) and polymerase having nuclease

PT activity useful in PCR
 PS Disclosure: Example 2: 75pp; English.
 CC The specificity of labelled probe release was examined by
 CC performing a PCR amplification using bacteriophage lambda DNA and
 CC primers, and a series of non-complementary kinased probes. The
 CC region to be amplified was a 500 base region on bacteriophage DNA
 CC from the GeneAmp DNA Amplification reagent kit, flanked by primers
 CC PCR01 (Q21948) and PCR02. Aliquots of labelled probes BW1,33 and
 CC BW35 (Q21945-47) were used, all of which were entirely non-
 CC complementary to the target sequence. On examination of the
 CC reaction prods. no probe degradation was found, confirming that
 CC the probe must be specifically bound to template before any
 CC degradation can occur.
 CC See also Q21943-66.
 SQ Sequence 25 BP; 7 A; 8 C; 6 G; 4 T;

alignment_scores:
 Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x Q21949 ..
 Align seg 1/1 to: Q21949 from: 1 to: 25

55 LeuSerLysserAlaThr 60
 |||||
 3 TTATCGAAATCAGCCACA 20

seq_name: N_Geneseq_36:Q24972

seq_documentation_block:

ID Q24972 standard; DNA; 25 BP.

AC Q24972;
 DT 24-NOV-1992 (first entry)
 DE PCR primer used to demonstrate a method of detecting target DNA.
 KW Amplification; marker; fluorescent pigment; phage lambda; ss.
 OS Synthetic.
 PN EP-487218-A.
 PD 27-MAY-1992.
 PF 31-OCT-1991; 310062.
 PR 31-OCT-1990; JP-294305.
 PA (TOYO) TOSOH CORP.
 PI Mitoma Y;
 DR WPI; 92-176819/22.
 PT Detecting or quantifying target nucleic acid in sealed vessel -
 PT using polymerase chain reaction and fluorescent pigment, to avoid
 PT sample contamination

FS Example 1; Page 6; 14pp; English.
 CC The PCR primer sequence is complementary to the + chain residues
 CC 7606-7630 of lambda DNA. The primer is used in a PCR reaction in
 CC conjunction with a fluorescent pigment whose fluorescence is changed
 CC upon reaction with nucleic acid. Since the measurement of the
 CC fluorescence intensity of the pigment can be carried out without
 CC removing the reaction mixt. after the PCR from the reaction vessel,
 CC the possibility of contamination by other samples can be reduced and
 CC the initial concn. of target nucleic acid may be determined
 CC accurately. The target nucleic acid to be amplified may be from a
 CC bacterium, virus or tissue, etc.
 CC See also Q24971-4.
 SQ Sequence 25 BP; 7 A; 8 C; 6 G; 4 T;

alignment_scores:
 Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x Q24972 ..

Align seg 1/1 to: Q24972 from: 1 to: 25

55 LeuSerLysserAlaThr 60
 |||||
 3 TTATCGAAATCAGCCACA 20

seq_name: N_Geneseq_36:Q38942

seq_documentation_block:

ID Q38942 standard; DNA; 25 BP.
 AC Q38942;
 DT 22-JUL-1993 (first entry)
 DE Sequencing primer for sequencing bacteriophage lambda DNA.
 KW Polymerase chain reaction; phage lambda template;
 KW primer extension; chain termination; DNA cycle sequencing; ss.
 OS Synthetic.
 PN WO9306243-A.
 PD 01-APR-1993.
 PF 28-AUG-1992; U07303.
 PR 27-SEP-1991; US-767137.
 PA (USBI-) US BIOCHEMICAL CORP.
 PI Fuller CW;
 DR WPI; 93-117559/14.
 PT DNA cycle sequencing method - by incorporating labelled
 PT nucleotide into extending DNA mol. during chain termination
 PS Example 1; Page 9; 36pp; English.

CC The primer corresponds to bases 7630-7606 in bacteriophage lambda.
 CC It was used in a cycle sequencing reaction with purified DNA from
 CC bacteriophage lambda as template. The primer is added in excess with
 CC delta-tag DNA polymerase, dGTP, dTTP and alpha-35S-labelled-dCTP. The
 CC primer is extended (until dATP is required) and labelled in a
 CC thermal cycler. The labelled primers can then be used in standard
 CC chain termination reactions for sequencing the template.
 SQ Sequence 25 BP; 7 A; 8 C; 6 G; 4 T;

alignment_scores:

Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x Q38942 ..
 Align seg 1/1 to: Q38942 from: 1 to: 25

55 LeuSerLysserAlaThr 60
 |||||
 3 TTATCGAAATCAGCCACA 20

seq_name: N_Geneseq_36:Q72900

seq_documentation_block:

ID Q72900 standard; DNA; 25 BP.
 AC Q72900;
 DT 03-JUN-1995 (first entry)
 DE PCR antisense primer for the amplification of bacteriophage lambda.
 KW PCR primer; bacteriophage lambda; ss.
 OS Synthetic.
 PN WO9424312-A.
 PD 27-OCT-1994.
 PF 19-APR-1994; U04275.
 PR 19-APR-1993; US-050681.
 PA (BECI) BECKMAN INSTR INC.
 PI Coassin PJ, Cook RM, Konrad KD, Rampal JB;
 DR WPI; 94-341888/42.

PT Method for the amplification of double stranded nucleic acid
 PT macromolecules - using inverse linkage oligo:nucleotide(s) to
 PT reduce non-specific hybridisation.(Eng).
 PS Example; Page 59; 95pp; English.
 CC Amplification of nucleotides 7131 to 7630 of bacteriophage lambda
 CC target segment was accomplished following PCR protocols using a
 CC Perkin Elmer CetusGeneAmp DNA Amplification Reagent Kit with

CC AmpliTag DNA Polymerase (Part No. N801-0055). The GeneAmp kit
CC includes the primers with sequences in Q72899 and Q72900 for
CC amplification.
SQ Sequence 25 BP; 7 A; 8 C; 6 G; 4 T;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x Q72900 ..
Align seg 1/1 to: Q72900 from: 1 to: 25

55 LeuserLysSerAlaThr 60
|||||
3 TTATCGAATCAGCCACA 20

seq_name: N_Geneseq_36:Q72902

seq_documentation_block:

ID Q72902 standard; DNA; 25 BP.

DT 03-JUN-1995 (first entry)

DE 5'-3' sequence of inverse linkage oligonucleotide (ILO) A-180.
KW Inverse linkage oligonucleotide; ILO; ss.

OS Synthetic.

FH Key Location/Qualifiers

FT misc_feature 1

FT /*tag= a
/label= inverse linkage

FT WO9424312-A.

PD 27-OCT-1994.

PF 19-APR-1994; U04275.

PR 19-APR-1993; US-050681.

PA (BECI) BECKMAN INSTR INC.

PI Coassin PJ, Cook RM, Konrad KD, Rampal JB;

DR WPI; 94-34188/42.

PT Method for the amplification of double stranded nucleic acid
PT macromolecules - using inverse linkage oligo:nucleotide(s) to
PT reduce non-specific hybridisation.(Eng).

PS Example; Page 63; 95pp; English.

CC An Inverse Linkage Oligonucleotide (ILO) has at least two

CC 3' termini or at least two 5' termini. The Perkin Elmer CetusGeneAmp

CC DNA Amplification Reagent Kit with AmpliTag DNA Polymerase (Part No.

CC N801-0055) includes the primers with sequences in Q72899

CC and Q72900 for amplification. ILOs were synthesised utilising both

CC the sense and the antisense primers from the GeneAmp kit. In ILO

CC A-180, the sequence is 3'-5' Q72901 - Inverse Linkage - 5'-3',

CC Q72902.

SQ Sequence 25 BP; 7 A; 8 C; 6 G; 4 T;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x Q72902 ..
Align seg 1/1 to: Q72902 from: 1 to: 25

55 LeuserLysSerAlaThr 60
|||||
3 TTATCGAATCAGCCACA 20

seq_name: N_Geneseq_36:Q72903

seq_documentation_block:

ID Q72903 standard; DNA; 25 BP.

AC Q72903;
DT 03-JUN-1995 (first entry)
DE 3'-5' sequence of inverse linkage oligonucleotide (ILO) A-181.
KW Inverse linkage oligonucleotide; ILO; ss.

OS Synthetic.

FH Key Location/Qualifiers

FT misc_feature 25

FT /*tag= a
/label= inverse linkage

FT WO9424312-A.

PD 27-OCT-1994.

PF 19-APR-1994; U04275.

PR 19-APR-1993; US-050681.

PA (BECI) BECKMAN INSTR INC.

PI Coassin PJ, Cook RM, Konrad KD, Rampal JB;

DR WPI; 94-34188/42.

PT Method for the amplification of double stranded nucleic acid
PT macromolecules - using inverse linkage oligo:nucleotide(s) to
PT reduce non-specific hybridisation.(Eng).

PS Example; Page 64; 95pp; English.

CC An Inverse Linkage Oligonucleotide (ILO) has at least two

CC 3' termini or at least two 5' termini. The Perkin Elmer CetusGeneAmp

CC DNA Amplification Reagent Kit with AmpliTag DNA Polymerase (Part No.

CC N801-0055) includes the primers with sequences in Q72899

CC and Q72900 for amplification. ILOs were synthesised utilising both

CC the sense and the antisense primers from the GeneAmp kit. In ILO

CC A-181, the sequence is 3'-5' Q72903 - Inverse Linkage - 5'-3',

CC Q72904.

SQ Sequence 25 BP; 7 A; 8 C; 6 G; 4 T;

alignment_scores:

Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x Q72903 ..

Align seg 1/1 to: Q72903 from: 1 to: 25

55 LeuserLysSerAlaThr 60
|||||
3 TTATCGAATCAGCCACA 20

seq_name: N_Geneseq_36:Q82728

seq_documentation_block:

ID Q82728 standard; DNA; 25 BP.

AC Q82728;

DT 09-OCT-1995 (first entry)

DE M13mp18 phage DNA primer.

KW M13mp18 phage DNA; viral infection; hereditary diseases;

KW detection assay; primer; ss.

OS Synthetic.

PN EP-639647-A.

PD 22-FEB-1995.

PF 06-JUL-1994; 110526.

PR 08-JUL-1993; JP-168895.

PA (EIKE) EIKEN CHEM CO LTD.

PA (TANA) TANABE SEIYAKU CO.

PA (EIKE) EIKEN KAGAKU KK.

PI Shibata T, Umemura I, Yamagata K;

DR WPI; 95-083458/12.

PT Assay for detecting a target nucleic acid sequence - using a pair
PT of primers, a fluorescent-labelled probe and a nucleic acid

PT polymerase having exo:nuclease activity

PS Example 1; Page 11; 12pp; English.

CC Q82724 is a fluorescent-labelled probe for M13mp18 phage DNA, it

CC was used to demonstrate a new assay for the detection of a target

CC nucleic acid sequence. The assay also used the primer pairs

CC Q82725/26 and Q82727/28, and a nucleic acid polymerase with

CC endonuclease activity (pref. from Thermus aquaticus). The assay

CC could be used for the detection of hereditary diseases and viral
 CC infections.
 SQ Sequence 25 BP; 7 A; 8 C; 6 G; 4 T;

alignment_scores:
 Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x Q82728 ..
 Align seg 1/1 to: Q82728 from: 1 to: 25

55 LeuSerLysSerAlaThr 60
 |||||
 3 TTATCGAAATCAGCCACA 20

seq_name: N_Geneseq_36:T12351
 seq_documentation_block:
 ID T12351 standard; DNA; 25 BP.
 AC T12351;
 DT 11-SEP-1996 (first entry)
 DE Primer PCR02 for bacteriophage lambda fragment.
 KW Polymerase chain reaction; primer; PCR; amplif; bacteriophage; nuclease;
 KW polymerase; probe; pathogen; HIV; HTLV; ss.
 OS Synthetic.
 PN US5487972-A.
 PD 30-JAN-1996.
 PF 06-AUG-1990; 563758.
 PR 06-AUG-1990; US-563758.
 PR 05-JAN-1993; US-961884.
 PA (HOFF) HOFFMANN LA ROCHE INC.
 PI Gelfand DH, Holland PM, Saiki RK, Watson RM;
 DR WPI: 96-105223/11
 PT Assay for target nucleic acid sequences - by amplification using a
 PT template-dependent nucleic acid polymerase having 5' to 3' nuclease
 PT activity.
 PS Example 2; Column 17; 44pp; English.
 CC T12350 and T12351 represent amplification primers for a 500 bp region of
 CC bacteriophage lambda. This sequence corresponds to nucleotides T131-7155
 CC of the bacteriophage sequence. The amplified sequence was used to test
 CC the specificity of the probes represented by T12347-T12349. The probes
 CC are used in the process of the invention, which is for the detection or
 CC measurement of a target nucleic acid (TNA) sequence in a sample. The TNA
 CC is amplified by a primer set (such as these two sequences), using a
 CC nucleic acid polymerase which has 5' to 3' nuclease activity. Contained
 CC within the reaction mixture is a labelled oligonucleotide probe which is
 CC complementary to a region of the TNA. The probe used is complementary to
 CC a portion of the TNA adjacent to the 3' end of one of the probes, and
 CC contains a tail of nucleotides that are not complementary to the TNA
 CC sequence. The release of labelled fragments is then measured or detected
 CC to determine the presence or absence of the TNA in the sample. This
 CC method can be used for the detection of pathogen DNA (such as HIV or HTLV
 CC DNA) in a sample. The method also allows for both amplification of a
 CC target and the release of a label for detection to be accomplished in a
 CC reaction system without resorting to multiple handling steps.
 SQ Sequence 25 BP; 7 A; 8 C; 6 G; 4 T;

alignment_scores:
 Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x T12351 ..
 Align seg 1/1 to: T12351 from: 1 to: 25

55 LeuSerLysSerAlaThr 60

|||||
 3 TTATCGAAATCAGCCACA 20

seq_name: N_Geneseq_36:V21785
 seq_documentation_block:
 ID V21785 standard; DNA; 25 BP.
 AC V21785;
 DT 14-JUL-1998 (first entry)
 DE Bacteriophage lambda specific sequencing primer.
 KW DNA sequencing; bacteriophage lambda; DNA polymerase; extended primer;
 KW primer; ss.
 OS Synthetic.
 OS Bacteriophage lambda.
 PN US5741640-A.
 PD 21-APR-1998.
 PF 18-MAY-1995; 443468.
 PR 14-APR-1994; US-227702.
 PR 27-SEP-1991; US-767137.
 PR 18-MAY-1995; US-443468.
 PA (AMSH) AMERSHAM LIFE SCI INC.
 PI Fuller CW;
 DR WPI: 98-260495/23.
 PT Method for sequencing DNA, e.g. plasmid DNA or lambda DNA -
 PT comprising contacting polynucleotide primer in presence of DNA
 PT polymerase to form extended primer, dissociating extended primer,
 PT and contacting extended primer again
 PS Example 5; Column 8; 15pp; English.
 CC This primer corresponds to bases 7630-7606 of bacteriophage lambda and
 CC can be used in a method for sequencing DNA. The method comprises
 CC contacting a polynucleotide primer complementary to a region of the
 CC DNA to be sequenced, the DNA to be sequenced together in the presence of
 CC a DNA polymerase, and 1-3 dNTP's, at least one of which is labelled to
 CC allow extension of the primer by addition of one or more of the dNTP's
 CC to the primer, to form an extended primer. The extended primer is
 CC dissociated from the DNA and the contacting and dissociating steps are
 CC repeated a number of times. The extended primer is contacted with the DNA
 CC to be sequenced in the presence of a DNA polymerase, 4 dNTP's, and a
 CC chain terminating agent. The process does not require the preparation of
 CC a labelled primer, and requires the same apparatus as thermal cycle
 CC sequencing, so the two methods can be conveniently used together. The
 CC process also allows the use of labels which are not otherwise usable with
 CC polynucleotide kinase, such as alpha-35S dNTP's, fluorescent
 CC nucleotides, and nucleotides labelled with other useful groups (e.g.
 CC biotin). Repeated cycling of the labelling step using limited amounts of
 CC template, increases the amount of product to a level sufficient for
 CC normal detection procedures, such as autoradiography with overnight
 CC exposure. Furthermore, because the process allows sequencing of small
 CC quantities of template, the need to grow clones and purify DNA from them
 CC is unnecessary. The chance of adding impurities into the sequencing
 CC reaction along with the template, is also reduced.
 SQ Sequence 25 BP; 7 A; 8 C; 6 G; 4 T;

alignment_scores:
 Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x V21785 ..
 Align seg 1/1 to: V21785 from: 1 to: 25

55 LeuSerLysSerAlaThr 60
 |||||
 3 TTATCGAAATCAGCCACA 20

seq_name: N_Geneseq_36:V21425
 seq_documentation_block:
 ID V21425 standard; DNA; 25 BP.
 AC V21425;

DT 04-AUG-1998 (first entry)
DE Bacteriophage lambda genome primer 2.
KW Primer; gene mapping; RecA-like recombinase; human p53; lambda phage;
KW in situ hybridisation; infectious disease; gene therapy; PCR;
KW amplification; ss.
OS Synthetic.
PN Bacteriophage lambda.
PN WO9808975-A1.
PD 05-MAR-1998.
PD 29-AUG-1997; J03019.
PR 26-DEC-1996; JP-347090.
PR 29-AUG-1996; JP-229061.
PA (DAIK) DAIKIN IND LTD.
PI Kigawa K, Kusumi K, Mukai E, Obata K, Yamanaka M;
DR WPI; 98-179448/16.
PT Targeting DNA by reaction with homologous probe and recombinase -
PT with addition of heterologous probe to increase efficiency and
PT sensitivity, useful for in situ hybridisation, gene therapy, cloning
PT of genes, gene mapping etc.
PS Example 1; Page 22; 59pp; English.
CC Primers 2 and 1 (V21424) were used to amplify the Bacteriophage lambda
CC heterologous probe 1 (V21409). The invention provides a method for
CC detecting target DNA in a sample using RecA-like recombinase and probes
CC homologous to the target DNA. Addition of heterologous probes was found
CC to increase the sensitivity and efficiency of detecting the target DNA.
CC In the example given, the target DNA was the human p53 cDNA which was
CC incorporated in a pHP53B plasmid. Homologous probes 1-6 (V21407, V21408,
CC V21412, V21414-V21416) containing a partial sequence of the p53 cDNA were
CC used. The heterologous probes 1-4 (V21409-V21411 and V21413)
CC corresponded to a partial sequence of the lambda phage genome. To
CC measure the specificity of detecting the target DNA, a heterologous
CC double stranded circular DNA (plasmid vector pUC18) was also used. The
CC method is useful for detecting the target DNA in cells by in situ
CC hybridisation and to target the DNA in living cells by in vivo gene
CC targeting, e.g. for gene therapy. The method is also claimed to be
CC useful for isolating and cloning target genes from libraries, in gene
CC mapping and for detecting genetic aberrations or mutations or infectious
CC diseases.
SQ Sequence 25 BP; 7 A; 8 C; 6 G; 4 T;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x V21425 ..
Align seg 1/1 to: V21425 from: 1 to: 25

seq_name: N_Geneseq_36:X15325

seq_documentation_block:
ID X15325 standard; DNA; 25 BP.
AC X15325;
DT 29-APR-1999 (first entry)
DE Oligonucleotide of the invention.
KW Thermostable polypeptide factor; DNA synthesis activity;
KW DNA polymerase; in vitro DNA synthesis; ss.
OS Synthetic.
PN WO9900506-A1.
PD 07-JAN-1999.
PD 24-JUN-1998; J02845.
PR 21-NOV-1997; JP-320692.
PR 26-JUN-1997; JP-187496.
PA (TAKI) TAKARA SHUZO CO LTD.
PI Asada K, Fujita T, Kato I, Miyake K, Mukai H, Sato Y,
PI Umori T;

DR WPI; 99-095751/08.
PT Thermostable polypeptide factors promoting the activity of DNA
PT polymerase - for improvement of DNA synthesis and amplification in
PT vitro.
PS Example 20; Page 150; 177pp; Japanese.
CC The specification describes Pyrococcus furiosus thermostable polypeptide
CC factors. These factors bind to, and promote the DNA synthesis activity
CC of DNA polymerase. The polymerase related factors can be used to
CC provide more efficient in vitro DNA synthesis and amplification
CC systems (e.g. for polymerase chain reaction) by using the factors in
CC conjunction with a DNA polymerase. The present oligonucleotide was
CC used in the course of the invention.
SQ Sequence 25 BP; 7 A; 8 C; 6 G; 4 T;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x X15325 ..
Align seg 1/1 to: X15325 from: 1 to: 25

seq_name: N_Geneseq_36:X25832

seq_documentation_block:
ID X25832 standard; DNA; 25 BP.
AC X25832;
DT 08-JUN-1999 (first entry)
DE Primer #2 for bacteriophage lambda sequence.
KW Analysis; mutation; binding reagent; chromatography; separation; genome;
KW detection; primer; PCR; amplification; bacteriophage; lambda; ss.
OS Synthetic.
OS Bacteriophage lambda.
PN WO9909203-A1.
PD 25-FEB-1999.
PD 18-AUG-1998; U17062.
PR 14-OCT-1997; US-062413.
PR 18-AUG-1997; US-055676.
PA (TRAN-) TRANSGENOMIC INC.
PI Gjerde DT, Taylor PD;
DR WPI; 99-190174/16.
PT Analysing sample of double stranded DNA - useful for determining
PT genetic mutations
PS Example 1; Page 31; 50pp; English.
CC The invention relates to a method for analysing a sample of double
CC stranded DNA to determine the presence of a mutation by:
CC (a) contacting the sample with a mutation site binding reagent; and
CC (b) chromatographically separating and detecting the product of (a).
CC Primers X25831-X25832 were used to PCR amplify a 500 bp fragment of
CC the bacteriophage lambda genome for use as a control sequence in the
CC method. The method is useful for the detection of mutations which
CC is important in the field of diagnosing diseases, understanding their
CC origins and the development of potential treatments.
SQ Sequence 25 BP; 7 A; 8 C; 6 G; 4 T;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x X25832 ..
Align seg 1/1 to: X25832 from: 1 to: 25

```

55 LeuSerLysSerAlaThr 60
|||||
3 TTATCGAAATCAGCCACA 20

seq_name: N_Geneseq_36:X19804

seq_documentation_block:
ID X19804 standard; DNA; 25 BP.
AC X19804;
DT 10-JUN-1999 (first entry)
DE PCR primer SEQ ID NO:3 from US5891682.
KW PCR primer; radiation; ultraviolet B radiation; dosimeter; exposure; ss.
KW high-energy particle radiation; X-radiation; dosimeter; exposure; ss.
OS Synthetic.
PN US5891682-A.
PD 06-APR-1999.
PF 06-MAY-1997; 851634.
PR 06-MAY-1996; US-016892.
PR 06-MAY-1997; US-851634.
PA (FLOR-) FLORIDA INST TECHNOLOGY.
PI Regan JD, Yoshida H;
DR WPI; 99-253859/21.
PT Dosimeter for measuring ultraviolet exposure
PS Disclosure; Fig 10; 24pp; English.
CC The present invention describes a dosimeter for measuring radiation
CC dosage in a predetermined ultraviolet (UV) wavelength range. The
CC dosimeter can be used to measure the radiation dosage incident on a
CC selected object or area. The dosimeter is used to measure exposure to
CC solar UV-B radiation (290-400 nm), particularly of a human or animal
CC and may also be used to measure dosage of X-rays and high energy
CC particle radiation. The dosimeter is compact (only a few ng of DNA are
CC needed), portable and inexpensive. It can be used under extreme
CC conditions, e.g. under water, and measures biologically active
CC radiation. The present sequence represents a PCR primer used in the
CC exemplification of the present invention.
SQ Sequence 25 BP; 7 A; 8 C; 6 G; 4 T;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522b-2_COPY_335_428 x X19804 ..
Align seg 1/1 to: X19804 from: 1 to: 25

55 LeuSerLysSerAlaThr 60
|||||
3 TTATCGAAATCAGCCACA 20

seq_name: N_Geneseq_36:X19802

seq_documentation_block:
ID X19802 standard; DNA; 26 BP.
AC X19802;
DT 10-JUN-1999 (first entry)
DE PCR primer SEQ ID NO:1 from US5891682.
KW PCR primer; radiation; ultraviolet B radiation; dosimeter; exposure; ss.
KW high-energy particle radiation; X-radiation; dosimeter; exposure; ss.
OS Synthetic.
PN US5891682-A.
PD 06-APR-1999.
PF 06-MAY-1997; 851634.
PR 06-MAY-1996; US-016892.
PR 06-MAY-1997; US-851634.
PA (FLOR-) FLORIDA INST TECHNOLOGY.
PI Regan JD, Yoshida H;
DR WPI; 99-253859/21.
PT Dosimeter for measuring ultraviolet exposure
PS Disclosure; Fig 10; 24pp; English.
CC The present invention describes a dosimeter for measuring radiation.

```

```

CC dosage in a predetermined ultraviolet (UV) wavelength range. The
CC dosimeter can be used to measure the radiation dosage incident on a
CC selected object or area. The dosimeter is used to measure exposure to
CC solar UV-B radiation (290-400 nm), particularly of a human or animal
CC and may also be used to measure dosage of X-rays and high energy
CC particle radiation. The dosimeter is compact (only a few ng of DNA are
CC needed), portable and inexpensive. It can be used under extreme
CC conditions, e.g. under water, and measures biologically active
CC radiation. The present sequence represents a PCR primer used in the
CC exemplification of the present invention.
SQ Sequence 26 BP; 7 A; 8 C; 6 G; 5 T;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522b-2_COPY_335_428 x X19802 ..
Align seg 1/1 to: X19802 from: 1 to: 26

55 LeuSerLysSerAlaThr 60
|||||
3 TTATCGAAATCAGCCACA 20

seq_name: N_Geneseq_36:V19100

seq_documentation_block:
ID V19100 standard; DNA; 30 BP.
AC V19100;
DT 13-AUG-1998 (first entry)
DE Bacillus amyloliquefaciens subtilisin gene PCR primer SEQ ID NO:19.
KW Eglin gene; fusion protein; protease; inhibitor; cleaning; detergent;
KW oligonucleotide-directed mutagenesis; PCR primer; laundry; SS;
KW hard-surface cleaner; dishwashing composition; oral hygiene composition;
KW denture cleaner; contact lens; Streptomyces subtilisin inhibitor; ss.
OS Synthetic.
OS Bacillus amyloliquefaciens.
PN WO9813483-A1.
PD 02-APR-1998.
PF 23-SEP-1997; U16354.
PR 24-SEP-1996; US-026947.
PA (PROC ) PROCTER & GAMBLE CO.
PI Armprilester JM, McIver JM, Saunders CW, Youngquist RS;
DR WPI; 98-230697/20.
PT Detergent compatible fusion protein - comprises protease and its
PT inhibitor linked by peptide chain, useful as, e.g. stable source of
PT protease in cleaning composition(s)
PS Example 4; Page 18; 81pp; English.
CC A new detergent-compatible fusion protein has been developed which
CC comprises: (i) detergent-compatible protease, and (ii) protein protease
CC inhibitor, linked by a peptide chain. The detergent-compatible fusion
CC protein is useful in cleaning compositions e.g. liquid or granular
CC laundry detergents or detergent additives, hard-surface cleaners,
CC dishwashing compositions, oral hygiene compositions, denture or contact
CC lens cleaners. The detergent-compatible fusion protein provides
CC increased yield and/or purity of the detergent-compatible protease, the
CC correct protease-protease inhibitor stoichiometry (contrast separate
CC expression of protease and protease inhibitor), and improve stability
CC of the protease (specifically against self-proteolysis) and so of
CC formulations. The present sequence represents a PCR primer from an
CC example of the present invention, for constructing a Bacillus
CC amyloliquefaciens subtilisin gene, used in the preparation of fusion
CC proteins.
SQ Sequence 30 BP; 4 A; 6 C; 7 G; 13 T;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:
US-09-332-522B-2_COPY_335_428 x V19100/rev ..
Align seg 1/1 to reverse of: V19100 from: 1 to: 30

52 GluAlaLysLeuSerLys 57
|||||
28 GAAGCAAACTGAGCAA 11

seq_name: N_Geneseq_36:V18004
seq_documentation_block:
ID V18004 standard; DNA; 30 BP.
AC V18004;
DT 13-AUG-1998 (first entry)
DE Primer used to amplify the subtilisin gene of B. amyloliquefaciens.
KW Streptomyces subtilisin inhibitor; SSI; detergent-stable;
KW protease inhibitor; cleaning composition; stabilise; protease;
KW subtilisin; PCR primer; ss.
OS Synthetic.
OS Bacillus amyloliquefaciens.
PN WO9813387-A1.
PD 02-APR-1998.
PF 23-SEP-1997; U16355.
PR 24-SEP-1996; US-026944.
PA (PROC ) PROCTER & GAMBLE CO.
PI Correa PE, Laskowski M, Saunders CW;
DR WPI; 98-230639/20.
PT Modified form of protease inhibitor stable to detergents and related
PT DNA - vectors, transformed cells and antibodies, releases protease
PT on dilution and useful in detergent and other cleaning compositions
PS Example 1; Page 12; 71pp; English.
CC PCR primers V18003-04 were used to amplify the subtilisin gene of
CC Bacillus amyloliquefaciens. The amplified product is cloned. The SSI
CC gene is excised from a vector and ligated into a second vector
CC containing a subtilisin gene. Mutants are produced using this construct.
CC The SSI protein is a detergent-stable protease inhibitor, and has a KI
CC value such that protease is bound in the cleaning composition but present
CC in free form during cleaning. The protein and its variants are used to
CC stabilise proteases in any type of cleaning composition (including
CC laundry detergents, hard surface cleaners, dishwashing compositions,
CC oral hygiene products and denture or contact lens cleaners), whether
CC solid, liquid or gaseous.
SQ Sequence 30 BP; 4 A; 6 C; 7 G; 13 T;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x V18004/rev ..
Align seg 1/1 to reverse of: V18004 from: 1 to: 30

52 GluAlaLysLeuSerLys 57
|||||
28 GAAGCAAACTGAGCAA 11

seq_name: N_Geneseq_36:V31624
seq_documentation_block:
ID V31624 standard; cDNA; 38 BP.
AC V31624;
DT 21-AUG-1998 (first entry)
DE Nucleotide sequence of sequence tagging site CH32 PCR primer 24.
KW Human; chromosome 8q12; sequence tagging site; STS; PLAG1; PLAG2;
KW CTNNB1; tumorigenesis gene; T-gene; antibody; benign tumour;
KW malignant tumour; leukaemia; lymphoma; cancer; inhibition; PCR;
KW amplification; primer; ss.
OS Synthetic.

```

```

OS Homo sapiens.
PN EP-825198-A1.
PD 25-FEB-1998.
PF 17-JAN-1997; 200130.
PR 22-AUG-1996; EP-202339.
PA (KULE-) KU LEUVEN RES & DEV.
PA (UYGO-) UNIV GOTEBOGRS HOLDINGBOLAGET AB.
PA (LEUV-) LEUVEN RES & DEV.
PI Kas KP, Stenman KGD, Van De Ven WJM, Voz ML;
DR WPI; 98-13252/13.
PT New tumorigenesis T-genes and proteins - useful for, e.g. preparing
PT antibodies for clinically diagnosing cells having non-physiological
PT proliferative capacity such as lipoblastomas
PS Example 1; Page 16; 71pp; English.
CC This is the nucleotide sequence of the PCR primer used for the
CC amplification of sequence tagging site (STS), in the method of the
CC invention, which involves isolation of the tumorigenesis genes
CC (T-gene), in the form of PLAG1, PLAG2, and CTNNB1 genes. Their
CC proteins can be used as a starting point for preparing antibodies for
CC clinically/medically diagnosing cells having a non-physiological
CC proliferative capacity as compared to wild type cells, where the former
CC cells are selected from both benign and malignant tumours, as well as
CC leukaemia and lymphomas. Derivatives of the T-gene are also used in
CC the diagnosis and preparation of therapeutical compositions for the
CC treatment of cancers, such as nucleic acid derivatives, and antibodies.
CC The T-gene may be used as a starting point for designing suitable
CC expression-modulating compounds or techniques for the treatment of
CC non-physiological proliferation phenomena in humans or animals.
CC Expression inhibitors of the T-gene can be used in the treatment of
CC diseases involving benign or malignant tumours.
SQ Sequence 38 BP; 8 A; 6 C; 12 G; 12 T;

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x V31624/rev ..
Align seg 1/1 to reverse of: V31624 from: 1 to: 38

57 LysSerAlaThrLeuArg 62
|||||
23 AAATCTGCAACACTGAGG 6

seq_name: N_Geneseq_36:T69438
seq_documentation_block:
ID T69438 standard; DNA; 40 BP.
AC T69438;
DT 29-JUL-1997 (first entry)
DE Plasmid p182SfII component oligonucleotide 10.
KW Oligonucleotide; plasmid; p182SfII; gene; promoter; R-TEM1;
KW beta-lactamase; construction; assembly; synthesis; end;
KW complementary; polymerase; reaction; ECPR; ss.
OS Synthetic.
PN WO9633207-A1.
PD 24-OCT-1996.
PF 18-APR-1996; U05480.
PR 18-APR-1995; US-425684.
PA (GLAX ) GLAXO GROUP LTD.
PI Lipshutz RJ, Stemmer WPC;
DR WPI; 96-485725/48.
PT Polynucleotide amplification using bivalent PCR primer to generate
PT rolling circle intermediate or linear concatemers - requires reduced
PT amt. of primer compared to conventional PCR
PS Example 2; Fig 5B; 77pp; English.
CC The present sequence is a component oligonucleotide of the 2.71 kb
CC plasmid p182SfII (Stemmer (1994) Nature 370:389), which encodes the
CC gene and promoter region for R-TEM1 beta-lactamase. A collection of
CC 134 component oligonucleotides, including the present sequence,

```

CC were synthesised and used to assemble the plasmid by end
 CC complementary polymerase reaction (ECPPr), employing the overlapping
 CC ends of the oligonucleotides.
 SQ Sequence 40 BP; 7 A; 7 C; 9 G; 17 T;

alignment_scores:
 Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x T69438 ..
 Align seg 1/1 to: T69438 from: 1 to: 40

45 LysValLeuLeuCysGly 50
 |||||
 5 AAAGTCTGCTATGTC 22

seq_name: N_Geneseq_36:T58933

seq_documentation_block:
 ID T58933 standard; DNA; 40 BP.
 AC T58933;
 DT 20-AUG-1997 (first entry)
 DE 3' PCR primer for linking human gastric lipase to signal peptide.
 KW Duodena; gastric; lipase; transgenic; plant; recombinant; extract; ss;
 KW food; absorption; fat; pancreatic; cystic fibrosis; exocrine; dairy; dog;
 KW hydrolysis; trans-esterification; substrate; enzyme; biofuel; human; PCR;
 KW polymerase chain reaction; amplification; primer; tobacco; leaf; seed.
 OS Synthetic.
 PN WO963277-A2.
 PD 24-OCT-1996.
 PF 19-APR-1996; F00605.
 PR 20-APR-1995; FR-004754.
 PA (BIOC-) BIOCEM SA.
 PA (LJOU) INST RECH JOUVEINAL.
 PI Baudino S, Benicourt C, Cudrey C, Gruber V, Lenee P;
 PI Merot B;
 PI WPI; 96-485783/48.
 DT Recombinant human or canine pre-duodenal lipase prodn. in transgenic
 PT plants - useful for facilitating absorption of fat, as bio-catalysts
 PT and for prodn. of bio-fuel
 PS Example II; Page 55; 130pp; French.
 CC The primers T58932-3 were used to PCR amplify the human gastric lipase
 CC gene (T58916) in which a BamHI restriction enzyme site is introduced at
 CC the 5' end in order to link the protein coding sequence with a signal
 CC peptide coding sequence. The signal peptide sequence is especially
 CC useful for expressing the recombinant lipase in a transgenic plant,
 CC especially in tobacco plant leaves or seeds.
 SQ Sequence 40 BP; 9 A; 8 C; 14 G; 9 T;

alignment_scores:
 Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x T58933/rev ..
 Align seg 1/1 to reverse of: T58933 from: 1 to: 40

8 SerGlnGlyThrIle 13
 |||||
 18 TCCAGGGCACCACCAT 1

seq_name: N_Geneseq_36:Q14788

seq_documentation_block:
 ID Q14788 standard; DNA; 41 BP.
 AC Q14788;

DT 24-FEB-1992 (first entry)
 DE Primer PRC02A-40 mer for the lambda viral genome.
 KW Homologous recombination; RecA; amplification; probe; RecA-803; ss.
 OS Synthetic.
 PN WO9117267-A.
 PD 14-NOV-1991.
 PF 17-APR-1991; U02626.
 PR 07-MAY-1990; US-520321.
 PA (STRI) SRI INTERNATIONAL.
 PI Zarling DA, Sena EP, Green CJ;
 PI WPI; 91-353786/48.
 DT Amplification of target DNA sequences - using primers, RecA
 PT protein, ATP-gamma-S, dNTPs and DNA polymerase.
 PS Example 7; Page 28; 52pp; English.
 CC The sequence was used to demonstrate RecA+ protein enhancement of
 CC DNA synthesis on native linear lambda DNA templates. The primer
 CC corresponds to nucleotides 7591-7630 of the lambda viral genome.
 CC It was coated in RecA+ protein using ATP-gamma-S. The coated
 CC primer was then added to 0.5 ug lambda viral genomic DNA SSB
 CC protein. After 3 mins at 37 deg.C, 1 U Klenow fragment was added.
 CC Results showed that the addition of RecA+ protein clearly enhanced
 CC bulk DNA synthesis.
 CC See also Q14781-88.
 SQ Sequence 41 BP; 9 A; 13 C; 8 G; 11 T;

alignment_scores:
 Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x Q14788 ..
 Align seg 1/1 to: Q14788 from: 1 to: 41

55 LeuSerIysSerAlaThr 60
 |||||
 3 TTATCGAAATCAGCCACC 20

seq_name: N_Geneseq_36:Q92983

seq_documentation_block:
 ID Q92983 standard; CDNA to mRNA; 46 BP.
 AC Q92983;
 DT 01-APR-1996 (first entry)
 DE Differentially expressed pre-invasive human breast marker gene DCIS-3.
 KW BRCA1; breast cancer; diagnosis; prognosis; gene therapy;
 KW non-comedo DCIS; ductal carcinoma in situ; intraductal carcinoma;
 KW pre-invasive human breast tissue; marker gene; ds.
 OS Homo sapiens sapiens.
 PN WO9519369-A1.
 PD 20-JUL-1995.
 PF 17-JAN-1995; U00608.
 PR 14-JAN-1994; US-182961.
 PR 17-JAN-1995; US-373799.
 PA (UYVA-) UNIV VANDERBILT.
 PI Holt JT, Jensen RA, Obermiller PS, Page DL, Robinson-Benion CL;
 PI Thompson ME;
 DR WPI; 95-269208/35.
 DT Detection, diagnosis and treatment of pre-invasive breast cancer -
 PT by identifying differentially expressed marker genes, also use of
 PT BRCA1 gene in therapy of breast cancer.
 PS Claim 35; Page 58; 149pp; English.
 CC Clone DCIS-3 was one of 10 differentially expressed cDNA clones
 CC in situ samples and from normal breast epithelial cells. DCIS-3
 CC is useful as a marker for pre-invasive human breast tissue. The
 CC present sequence is that of clone DCIS-3 as shown in the sequence
 CC listing of the patent; there are 3 additional nucleotides (i.e. AGT)
 CC between G9 and T10 in the DCIS-3 sequence shown in Figure 9 of the
 CC specification.
 SQ Sequence 46 BP; 8 A; 10 C; 14 G; 14 T;

alignment_scores:
 Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x Q92983/rev ..

Align seg 1/1 to reverse of: Q92983 from: 1 to: 46

55 LeuSerLysSerAlaThr 60

|||||

46 TTATCGAATACAGCCACA 29

seq_name: N_Geneseq_36:V15666

seq_documentation_block:

ID V15666 standard; DNA; 46 BP.

AC V15666;

DT 22-MAY-1998 (first entry)

DE PCR primer sequence.

KW Ligase detection reaction; LDR; PCR primer; ss.

OS Synthetic.

PN W09745559-Al.

PD 04-DEC-1997.

PF 27-MAY-1997; U09012.

PR 29-MAY-1996; US-018532.

PA (CORR) CORNELL RES FOUND INC.

PI Barany F, Belgrader P, Lubin M;

DR WPI; 98-032663/03.

PT Multiplex detection of nucleic acid sequence differences - using

ligase detection reaction coupled to PCR, useful for determining

gene dosage, for detecting genetic disorders, etc.

PS Example 3; Page 88; 158pp; English.

CC The present sequence was used in the development of three novel

methods for the detection nucleic acid sequence differences, i.e.

single-base changes, insertions, deletions or translocations. The

1st uses the ligase detection reaction (LDR) coupled to PCR, the

2nd a 1st PCR coupled to a 2nd PCR coupled to a LDR and the 3rd a

1st PCR coupled to a 2nd PCR.

Sequence 46 BP; 9 A; 15 C; 12 G; 10 T;

alignment_scores:
 Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x V15666/rev ..

Align seg 1/1 to reverse of: V15666 from: 1 to: 46

11 ThrThrIleIleValarg 16

|||||

30 ACCAGCATCATAGTCGG 13

seq_name: N_Geneseq_36:V20001

seq_documentation_block:

ID V20001 standard; DNA; 48 BP.

AC V20001;

DT 15-JUL-1998 (first entry)

DE PCR primer for mutated KGF-2 coding sequence.

KW Keratinocyte growth factor-2; KGF-2; fibroblast growth factor-12; FGF-12;

keratinocyte proliferation; growth stimulator; skin strength; mucositis;

epidermal thickening; wound healing; inflammatory bowel disease; therapy;

KW inflammation; hair growth promoter; PCR primer; ss.

OS Synthetic.

OS Homo sapiens.

PN W09806844-Al.
 PD 19-FEB-1998.
 PF 13-AUG-1997; U14112.
 PR 28-FEB-1997; US-039045.
 PR 13-AUG-1996; US-023852.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Coleman TA, Duan R, Jimenez P, Mendrick D, Moore PA,
 PI Ni J, Rampy MA, Ruben SM, Zhang J;
 DR WPI; 98-159536/14.
 PT Keratinocyte growth factor-2 deletion mutants - useful to promote
 or accelerate wound healing
 PS Example 12; Page 124; 251pp; English.
 CC This sequence is a primer for a mutated version of the human
 keratinocyte growth factor-2 (KGF-2) coding sequence of the invention.
 CC The KGF-2 protein is used to create the mutants of the invention, that
 CC stimulate proliferation of keratinocytes. The mutants have enhanced
 CC keratinocyte growth stimulating activity as compared to wild-type KGF-2
 CC (also known as fibroblast growth factor-12). They are used to stimulate
 CC growth or proliferation of keratinocytes. In particular, they are used to
 CC prevent or improve the appearance of wrinkles or aged skin, improving
 CC skin strength, promoting epidermal thickening, reducing scarring or
 CC improving wound healing, especially where an individual is wound healing
 CC impaired. Wounds to be treated may be surgical or excisional wounds, deep
 CC wounds involving damage of the dermis and epidermis, eye tissue wounds,
 CC dental tissue wounds, oral cavity wounds, diabetic, dermal, cubitus,
 CC arterial or venous stasis ulcers or burns. Treatment, especially by
 CC anastomosis, of wounds caused by colonic or gastrointestinal surgical
 CC procedures can also be achieved through use of the KGF-2 deletion
 CC mutants. They can also be used for treatment or prevention of mucositis,
 CC inflammatory bowel disease, reduction of inflammation, promoting hair
 CC growth, urothelial healing or tissue growth or repair in the female
 CC genital tract, or for treating tissue exposed to radiation or protecting
 CC tissue to be exposed to radiation (e.g. to allow an increase in radiation
 CC dosage used to treat a malignancy).
 CC Sequence 48 BP; 11 A; 13 C; 13 G; 11 T;

alignment_scores:
 Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x V20001/rev ..

Align seg 1/1 to reverse of: V20001 from: 1 to: 48

3 SerGlyGluThrMetSer 8

|||||

43 TCCGGGGAACCATGTCT 26

seq_name: N_Geneseq_36:Q23242

seq_documentation_block:

ID Q23242 standard; DNA; 49 BP.

AC Q23242;

DT 05-AUG-1992 (first entry)

DE Lambda bacteriophage PCR primer b.

KW Polymerase chain reaction; ss.

OS Synthetic.

PN US7683440-A.

PD 18-FEB-1992.

PF 10-APR-1991; 683440.

PR 10-APR-1991; US-683440.

PA (USSH) US DEPT HEALTH & HUMAN.

PI Shuldiner AR, Roth J;

DR WPI; 92-113928/14.

PT Ligase-free sub-cloning of polymerase chain reaction prod. - using

four primers and two rounds of PCR, is versatile, rapid and reliable

PS Example; Page 11; 29pp; English.

CC The sequence is that of a PCR primer used in the amplification of a

500 base pair fragment of lambda bacteriophage DNA. It is a 49-mer

CC that contains at its 3' end, a 25 nucleotide sequence complementary
 CC to the (+)-strand of lambda bacteriophage DNA, 500 bases downstream
 CC from where primer a (Q23241) would hybridise. The 5' end of the primer
 CC contg. the remaining 24 nucleotides (the 5' addition sequence) was
 CC identical to the other 3' end of the SmaI digested pGEM4Z. It is used
 CC as part of an example of a method for directionally subcloning PCR
 CC prods. into plasmid vectors without the use of DNA ligase. This
 CC method is faster than prior art methods and can be accomplished in
 CC a single day. It is also more versatile and reliable.
 CC See also Q23241-Q23246.
 SQ Sequence 49 BP; 13 A; 15 C; 11 G; 10 T;

alignment_scores:
 Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x Q23242 ..

Align seg 1/1 to: Q23242 from: 1 to: 49

55 LeuSerLysSerAlaThr 60
 |||||

27 TTATCGAAATCAGCCACA 44

seq_name: N_Geneseq_36:N91957

seq_documentation_block:

ID N91957 standard; DNA; 50 BP.

AC N91957; 1990 (first entry)

DE Complementary strand of the Neisseria gonorrhoeae Tem-1 beta lactamase

gene combined with 11a2c amplifier probe

Neisseria gonorrhoeae; beta lactamase Tem-1; capture probe;

amplifier probe 11a2c; temkit21.20; TEM-1NH assay.

OS Neisseria gonorrhoeae

FT Key Location/Qualifiers

FT misc_feature 1..30

FT /*tag= a

FT /*capture probe"

FT misc_feature 31..50

FT /*tag= b

FT /*11a2c amplifier probe"

PI W08903891-A.

PD 05-May-1989.

PF 14-OCT-1988; U03644.

PR 30-SEP-1988; US-252638, US-109282.

PA (CHIR-) Chiron Corp.

PI Urdea MS, Warner B, Running JA, Kolberg JA, Clyne JM, Sanchez-Pescador R;

DR WPI; 89-150787/20.

PT Nucleic acid multimer for hybridisation assays

PT - having single-stranded oligo-nucleotide units

PT capable of binding specifically to sequences of interest.

PS Fig. 10-2; 112pp; English.

CC Partial nucleotide sequences of the capture and amplifier probes used in

the TEM-1NH assay. The capture probe is complementary to the coding

strand of N. gonorrhoeae Tem-1 beta lactamase gene. It is called

temkit21.20.

SQ Sequence 50 BP; 14 A; 13 C; 12 G; 11 T;

alignment_scores:
 Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x N91957/rev ..

Align seg 1/1 to reverse of: N91957 from: 1 to: 50

45 LysValLeuLeuCysGly 50
 |||||

21 AAAGTCTGCTATGTCG 4

seq_name: N_Geneseq_36:Q61446

seq_documentation_block:

ID Q61446 standard; DNA; 50 BP.

AC Q61446; 1994 (first entry)

DE Bacteriophage lambda ligand DNA sequence.

KW Detection; bacteriophage lambda; differential hybridisation;

elution profile; separation column; ss.

OS Bacteriophage lambda.

FT Key Location/Qualifiers

FT modified_base 1

FT /*tag= a

FT /note= "joined to amino-link for immobilisation

on an insoluble carrier"

PN J06078799-A.

PD 22-MAR-1994.

PF 31-AUG-1992; 257148.

PR 31-AUG-1992; JP-257148.

PA (SUME) SUMITOMO ELECTRIC IND CO.

DR WPI; 94-131295/16.

PT Detecting specific DNA fragments - comprises immobilising ligand

on insol. carrier used as solid phase in elution column,

PT increasing the temp. and analysing elution profile

PS Example 1; Page 4; 6pp; Japanese.

CC A sequence (Q61446) from bacteriophage lambda was synthesised and

CC immobilised on silica gel via "aminolink" bound to its 5'-end. The

CC sequence is an example of a ligand to be bound to a target DNA

CC fragment. A DNA sample (made single-stranded by denaturation) is

CC contacted with the immobilised ligand which is packed into a column.

CC Only the target DNA is specifically bound and other DNA passes

CC through the column. The target DNA fragment is eluted by increasing

CC the column temperature.

SQ Sequence 50 BP; 10 A; 16 C; 12 G; 12 T;

alignment_scores:
 Quality: 6.00 Length: 6
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x Q61446 ..

Align seg 1/1 to: Q61446 from: 1 to: 50

55 LeuSerLysSerAlaThr 60
 |||||

3 TTATCGAAATCAGCCACA 20

OM of: US-09-332-522B-2_COPY_335_428 to: Issued_Patents_NA:* out_format : pfs

Date: Sep 27, 2000 2:48 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=framet_p2n.model -DEV=xlip
-Q=/cgn2_1/USPTO_spool/US09333522/runat_27092000_091221_6548/app_query.fasta_1.1355
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=oli.rni -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-XGAPOP=6.000 -XGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM-ext -INLEN=0
-MAXLEN=200000000 -USER=US09332522 -CGN1_1.92 -NCPU=6 -ICPU=3
-LOGLOG -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-332-522B-2_COPY_335_428

Query length: 94

Database: Issued_Patents_NA.*

Database sequences: 243080

Database length: 68777915

Search time (sec): 77.950000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

```
Sequence      Strd Orig      ZScore      EScore Len | Documentation      ..
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-533-669A-1 + 7.00 103.85 270.61 3134 | Patent
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-892-272-5 + 6.00 120.68 31.27 23 | Se
/cgn2_6/ptodata/2/ina/5C_COMB.seq:PCT-US91-02626-5 + 6.00 120.37 32.51 24 | Patent
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-050-681A-2 + 6.00 120.08 33.74 25 | Se
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-961-884A-7 + 6.00 120.08 33.74 25 | Se
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-767-137A-1 + 6.00 120.08 33.74 25 | Se
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-443-468-1 + 6.00 120.08 33.74 25 | Se
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-437-318-1 + 6.00 120.08 33.74 25 | Se
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-428-941-7 + 6.00 120.08 33.74 25 | Se
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-851-634-3 + 6.00 120.08 33.74 25 | Se
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-851-634-5 + 6.00 120.08 33.74 25 | Se
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-851-634-7 + 6.00 120.08 33.74 25 | Se
/cgn2_6/ptodata/2/ina/5C_COMB.seq:PCT-US94-04275-2 + 6.00 120.08 33.74 25 | Patent
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-851-634-1 + 6.00 120.08 33.74 25 | Se
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-851-634-9 + 6.00 119.80 34.97 26 | Se
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-425-684-40 + 6.00 116.74 51.81 40 | Se
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-675-502-40 + 6.00 116.74 51.81 40 | Se
/cgn2_6/ptodata/2/ina/5C_COMB.seq:PCT-US91-02626-7 + 6.00 116.56 52.99 41 | Patent
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-182-961B-3 - 6.00 115.75 58.86 46 | Se
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-864-473-50 - 6.00 115.75 58.86 46 | Se
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-023-082A-45 - 6.00 115.44 61.19 48 | Se
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-050-681A-3 + 6.00 115.15 63.51 50 | Se
/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCT-US94-04275-3 + 6.00 115.15 63.51 50 | Patent
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-221-662-1 + 6.00 114.35 70.43 56 | Se
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-221-662-3 + 6.00 111.90 96.41 79 | Se
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-023-082A-33 - 6.00 111.81 97.52 80 | Se
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-182-961B-5 + 6.00 111.46 101.96 84 | Se
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-009-265-38 + 6.00 110.97 108.59 90 | Se
/cgn2_6/ptodata/2/ina/5C_COMB.seq:PCT-US94-06456-18 + 6.00 110.37 117.36 98 | Patent
/cgn2_6/ptodata/2/ina/5C_COMB.seq:PCT-US94-06456-47 + 6.00 110.37 117.36 98 | Patent
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-648-796A-18 - 6.00 110.22 119.54 100 | Patent
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-751-304B-2 - 6.00 109.29 134.73 114 | Patent
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-633-879C-7 + 6.00 107.20 176.22 153 | Se
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-193-078B-19 + 6.00 103.73 274.81 249 | Se
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-171-385-9 - 6.00 103.05 299.88 274 | Se
```

```
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-361-441B-9 - 6.00 103.05 299.88 274 | Se
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-994-946A-4 - 6.00 102.41 335.74 300 | Se
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-171-385-11 - 6.00 102.20 334.64 309 | Se
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-361-441B-11 - 6.00 102.20 334.64 309 | Se
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-171-385-18 - 6.00 101.95 345.50 320 | Se
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-361-441B-18 - 6.00 101.95 345.50 320 | Se
```

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-533-669A-1

```
seq_documentation_block:
; Sequence 1, Application US/08533669A
; Patent No. 5834592
; GENERAL INFORMATION:
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; THERAPY AND DIAGNOSIS OF LEISHMANIASIS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,669A
; FILING DATE: 22-SEP-1995
; CLASSIFICATION: 424
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3134 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 421..2058
; US-08-533-669A-1
```

```
alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-533-669A-1 ..
Align seg 1/1 to: US-08-533-669A-1 from: 1 to: 3134
58 SerAlaThrLeuArgGala 64
|||||
902 TCCGGACCCCTCCGAGGCC 922
```

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-892-272-5

```
seq_documentation_block:
; Sequence 5, Application US/08892272
; Patent No. 5891687
; GENERAL INFORMATION:
; APPLICANT: SCHLIEPER, Daniel
```

```
;
; APPLICANT: SOBEK, Harald
; APPLICANT: SCHMIDT, Manfred
; APPLICANT: VON WILKEN-BERGEMANN, Brigitte
; APPLICANT: MULLER-HILL, Banno
; TITLE OF INVENTION: POSITIVE SELECTION VECTOR BASED ON THE
; TITLE OF INVENTION: CAPS GENE, PCAPS VECTOR AND ITS USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikaigo, Marmelstein, Murray & Oram
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/892,272
; FILING DATE: 14-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 30 617.5
; FILING DATE: 29-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MURRAY, Robert B.
; REGISTRATION NUMBER: 22,980
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
; US-08-892-272-5
;
; alignment_scores:
; Quality: 6.00 Length: 6
; Ratio: 1.000 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 100.000
;
; alignment_block:
; US-09-332-522B-2_COPY_335_428 x US-08-892-272-5
;
; Align seg 1/1 to: US-08-892-272-5 from: 1 to: 23
;
; 55 LeuserLysSerAlaThr 60
; |||||
; 3 TTATCGAAATCAGCCACA 20
;
; seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US91-02626-5
;
; seq_documentation_block:
; Sequence 5, Application PC/TUS9102626
; GENERAL INFORMATION:
; APPLICANT: Zarling, David A.
; APPLICANT: Sena, Elissa P.
; TITLE OF INVENTION: Process for Nucleic Acid Hybridization
; TITLE OF INVENTION: and Amplification
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Swiss
; STREET: P.O. Box 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
```

```
;
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/02626
; FILING DATE: 19910405
; CLASSIFICATION: 435.6
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/520,321
; FILING DATE: 07-MAY-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Dehlinger, Peter J.
; REGISTRATION NUMBER: 28,006
; REFERENCE/DOCKET NUMBER: P-2579/8255-001.41
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 323-8302
; TELEFAX: (415) 323-8306
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: PRIMER, PRCO2-25, LAMBDA SEQUENCE
;
; PCT-US91-02626-5
;
; alignment_scores:
; Quality: 6.00 Length: 6
; Ratio: 1.000 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 100.000
;
; alignment_block:
; US-09-332-522B-2_COPY_335_428 x PCT-US91-02626-5
;
; Align seg 1/1 to: PCT-US91-02626-5 from: 1 to: 24
;
; 55 LeuserLysSerAlaThr 60
; |||||
; 3 TTATCGAAATCAGCCACA 20
;
; seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5223414-4
;
; seq_documentation_block:
; Patent No. 5223414
; APPLICANT: ZARLING, DAVID A.; SENA, ELISSA P.; GREEN,
; CHRISTOPHER J.
; TITLE OF INVENTION: PROCESS FOR NUCLEIC ACID
; HYBRIDIZATION AND AMPLIFICATION
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/520,321
; FILING DATE: 07-MAY-1990
; SEQ ID NO:4:
; LENGTH: 24
; 5223414-4
;
; alignment_scores:
; Quality: 6.00 Length: 6
; Ratio: 1.000 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 100.000
;
; alignment_block:
; US-09-332-522B-2_COPY_335_428 x 5223414-4
;
; Align seg 1/1 to: 5223414-4 from: 1 to: 24
```

```
55 LeuSerLySserAlaThr 60
|||||
3 TTATCGAATCAGCCACA 20

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-050-681A-2

seq_documentation_block:
; Sequence 2, Application US/08050681A
; Patent No. 5462854
; GENERAL INFORMATION:
; APPLICANT: Coassin, Peter J.
; APPLICANT: Cook, Donald M.
; APPLICANT: Rampal, Jang B.
; APPLICANT: Konrad, Kenneth D.
; TITLE OF INVENTION: Inverse Linkage
; TITLE OF INVENTION: Oligonucleotides for Chemical
; TITLE OF INVENTION: and Enzymatic Processes
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Beckman Instruments, Inc.
; STREET: 2500 Harbor Boulevard
; CITY: Fullerton
; STATE: California
; COUNTRY: USA
; ZIP: 92634
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb
; COMPUTER: IBM
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,681A
; FILING DATE: 19930419
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Burgoon, Richard P.
; REGISTRATION NUMBER: 34,787
; REFERENCE/DOCKET NUMBER: 129D-1100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 773-7936
; TELEFAX: (714) 773-7936
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: no
; ANTI-SENSE: yes
; US-08-050-681A-2

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-050-681A-2 ..
Align seg 1/1 to: US-08-050-681A-2 from: 1 to: 25

55 LeuSerLySserAlaThr 60
|||||
3 TTATCGAATCAGCCACA 20

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-07-961-884A-7

seq_documentation_block:
; Sequence 7, Application US/07961884A
; Patent No. 5487972
```

```
; GENERAL INFORMATION:
; APPLICANT: David H. Gelfand, Pamela M. Holland, Randall K. Saiki,
; APPLICANT: and Robert M. Watson
; TITLE OF INVENTION: Homogeneous Assay System
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110-1199
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/961,884A
; FILING DATE: January 5, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 563,758
; FILING DATE: August 6, 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Stacey R. Sias
; REGISTRATION NUMBER: 32,630
; REFERENCE/DOCKET NUMBER: 8737
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2863
; TELEFAX: (510) 814-2977
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other Nucleic Acid
; US-07-961-884A-7

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-07-961-884A-7 ..
Align seg 1/1 to: US-07-961-884A-7 from: 1 to: 25

55 LeuSerLySserAlaThr 60
|||||
3 TTATCGAATCAGCCACA 20

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-07-767-137A-1

seq_documentation_block:
; Sequence 1, Application US/07767137A
; Patent No. 5674679
; GENERAL INFORMATION:
; APPLICANT: Carl W. Fuller
; TITLE OF INVENTION: DNA CYCLE SEQUENCING
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/767.137A
; FILING DATE: September 27, 1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application(s) none
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 194/143
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-767-137A-1

alignment_scores:
    Quality: 6.00      Length: 6
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-07-767-137A-1 ..
Align seg 1/1 to: US-07-767-137A-1 from: 1 to: 25

55 LeuserLysserAlathr 60
|||||
3 TTATCGAAATCAGCCACA 20

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-443-468-1

seq_documentation_block:
; Sequence 1, Application US/08443468
; Patent No. 5741640
; GENERAL INFORMATION:
; APPLICANT: Carl W. Fuller
; TITLE OF INVENTION: DNA CYCLE SEQUENCING
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; COUNTRY: California
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443.468
; FILING DATE: May 18, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 07/767,137
; FILING DATE: September 27, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 213/213
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-443-468-1

alignment_scores:
    Quality: 6.00      Length: 6
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-443-468-1 ..
Align seg 1/1 to: US-08-443-468-1 from: 1 to: 25

55 LeuserLysserAlathr 60
|||||
3 TTATCGAAATCAGCCACA 20

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-437-318-1

seq_documentation_block:
; Sequence 1, Application US/08437318
; Patent No. 5741676
; GENERAL INFORMATION:
; APPLICANT: Carl W. Fuller
; TITLE OF INVENTION: DNA CYCLE SEQUENCING
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/437,318
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 07/767,137
; FILING DATE: 21-SEPT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 213/060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
```

```
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-437-318-1

alignment_scores:
  Quality: 6.00      Length: 6
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
  US-09-332-522B-2_COPY_335_428 x US-08-437-318-1 ..
Align seg 1/1 to: US-08-437-318-1 from: 1 to: 25

55 LeuSerLysSerAlaThr 60
|||||
3 TTATCGAAATCAGCCACA 20

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-428-941-7

seq_documentation_block:
; Sequence 7, Application US/08428941
; Patent No. 5804375
; GENERAL INFORMATION:
; APPLICANT: David H. Gelfand, Pamela M. Holland, Randall K.
; APPLICANT: Saiki, and Robert M. Watson
; TITLE OF INVENTION: Homogeneous Assay System
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110-1199
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/428,941
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/961,884
; FILING DATE: January 5, 1993
; REFERENCE/DOCKET NUMBER: 8737
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2863
; TELEFAX: (510) 814-2977
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other Nucleic Acid
; US-08-428-941-7

alignment_scores:
  Quality: 6.00      Length: 6
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
  US-09-332-522B-2_COPY_335_428 x US-08-428-941-7 ..
Align seg 1/1 to: US-08-428-941-7 from: 1 to: 25

55 LeuSerLysSerAlaThr 60
|||||
3 TTATCGAAATCAGCCACA 20

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-851-634-3

seq_documentation_block:
; Sequence 3, Application US/08851634
; Patent No. 5891682
; GENERAL INFORMATION:
; APPLICANT: YOSHIDA, Hiroko
; APPLICANT: REGAN, JAMES D.
; TITLE OF INVENTION: RADIATION DOSIMETER AND METHOD OF MAKING AND USING
; FILE REFERENCE: 18623BB
; CURRENT APPLICATION NUMBER: US/08/851,634
; CURRENT FILING DATE: 1997-05-06
; EARLIER APPLICATION NUMBER: 60/016,892
; EARLIER FILING DATE: 1996-05-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:500 bp
; OTHER INFORMATION: (7131-7630) Primer 1
; US-08-851-634-3

alignment_scores:
  Quality: 6.00      Length: 6
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
  US-09-332-522B-2_COPY_335_428 x US-08-851-634-3 ..
Align seg 1/1 to: US-08-851-634-3 from: 1 to: 25

55 LeuSerLysSerAlaThr 60
|||||
3 TTATCGAAATCAGCCACA 20

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-851-634-5

seq_documentation_block:
; Sequence 5, Application US/08851634
; Patent No. 5891682
; GENERAL INFORMATION:
; APPLICANT: YOSHIDA, Hiroko
; APPLICANT: REGAN, JAMES D.
; TITLE OF INVENTION: RADIATION DOSIMETER AND METHOD OF MAKING AND USING
; FILE REFERENCE: 18623BB
; CURRENT APPLICATION NUMBER: US/08/851,634
; CURRENT FILING DATE: 1997-05-06
; EARLIER APPLICATION NUMBER: 60/016,892
; EARLIER FILING DATE: 1996-05-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
```

; OTHER INFORMATION: Description of Artificial Sequence:1080 bp
; OTHER INFORMATION: (6551-7630) Primer 4
US-08-851-634-5

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Smilarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-851-634-5 ..
Align seg 1/1 to: US-08-851-634-5 from: 1 to: 25

55 LeuSerLysSerAlaThr 60
|||||
3 TTATCGAAATCAGCCACA 20

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-851-634-7

seq_documentation_block:
; Sequence 7, Application US/08851634
; Patent No. 5891682
; GENERAL INFORMATION:
; APPLICANT: YOSHIDA, Hiroko
; APPLICANT: REGAN, JAMES D.
; TITLE OF INVENTION: RADIATION DOSIMETER AND METHOD OF MAKING AND USING
; FILE REFERENCE: 18623BB
; CURRENT APPLICATION NUMBER: US/08/851,634
; CURRENT FILING DATE: 1997-05-06
; EARLIER APPLICATION NUMBER: 60/016,892
; EARLIER FILING DATE: 1996-05-06
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:2240 bp
; OTHER INFORMATION: (5391-7630) Primer 3
US-08-851-634-7

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Smilarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-851-634-7 ..
Align seg 1/1 to: US-08-851-634-7 from: 1 to: 25

55 LeuSerLysSerAlaThr 60
|||||
3 TTATCGAAATCAGCCACA 20

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-04275-2

seq_documentation_block:
; Sequence 2, Application PC/TUS9404275
; GENERAL INFORMATION:
; APPLICANT: Coassin, Peter J.
; APPLICANT: Cook, Donald M.
; APPLICANT: Rampal, Jang B.
; APPLICANT: Konrad, Kenneth D.
; TITLE OF INVENTION: Inverse Linkage
; TITLE OF INVENTION: Oligonucleotides for Chemical
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Beckman Instruments, Inc.
; STREET: 2500 Harbor Boulevard
; CITY: Fullerton
; STATE: California
; COUNTRY: USA
; ZIP: 92634
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb
; COMPUTER: IBM
; OPERATING SYSTEM: MS.DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04275
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Burgoon, Richard P.
; REGISTRATION NUMBER: 34,787
; REFERENCE/DOCKET NUMBER: 129D-1100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 773-7610
; TELEFAX: (714) 773-7936
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: no
; ANTI-SENSE: yes
PCT-US94-04275-2

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Smilarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x PCT-US94-04275-2 ..
Align seg 1/1 to: PCT-US94-04275-2 from: 1 to: 25

55 LeuSerLysSerAlaThr 60
|||||
3 TTATCGAAATCAGCGACA 20

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5210015-7

seq_documentation_block:
; Patent No. 5210015
; APPLICANT: GELFAND, DAVID H.; HOLLAND, PAMELA M.; SAIKI,
; RANDALL K.; WATSON, ROBERT M.
; TITLE OF INVENTION: HOMOGENEOUS ASSAY SYSTEM USING THE
; NUCLEASE ACTIVITY OF A NUCLEIC ACID POLYMERASE
; NUMBER OF SEQUENCES: 22
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/563,758
; FILING DATE: 06-AUG-1990
; SEQ ID NO:7
; LENGTH: 25
5210015-7

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Smilarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x 5210015-7 ..

Align seg 1/1 to: 5210015-7 from: 1 to: 25

55 LeuSerLysSerAlaThr 60
|||||
3 TTATCGAAATCAGCCACA 20

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-851-634-1

seq_documentation_block:

; Sequence 1, Application US/08851634

; Patent No. 5891682

; GENERAL INFORMATION:

; APPLICANT: YOSHIDA, Hiroko

; APPLICANT: REGAN, JAMES D.

; TITLE OF INVENTION: RADIATION DOSIMETER AND METHOD OF MAKING AND USING

; FILE REFERENCE: 18623BB

; CURRENT APPLICATION NUMBER: US/08/851.634

; CURRENT FILING DATE: 1997-05-06

; EARLIER APPLICATION NUMBER: 60/016.892

; EARLIER FILING DATE: 1996-05-06

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 26

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:230 BP

; OTHER INFORMATION: (7401-7630) Primer 0

US-08-851-634-1

alignment_scores:

Quality: 6.00 Length: 6

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x US-08-851-634-1 ..

Align seg 1/1 to: US-08-851-634-1 from: 1 to: 26

55 LeuSerLysSerAlaThr 60
|||||
3 TTATCGAAATCAGCCACA 20

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-851-634-9

seq_documentation_block:

; Sequence 9, Application US/08851634

; Patent No. 5891682

; GENERAL INFORMATION:

; APPLICANT: YOSHIDA, Hiroko

; APPLICANT: REGAN, JAMES D.

; TITLE OF INVENTION: RADIATION DOSIMETER AND METHOD OF MAKING AND USING

; FILE REFERENCE: 18623BB

; CURRENT APPLICATION NUMBER: US/08/851.634

; CURRENT FILING DATE: 1997-05-06

; EARLIER APPLICATION NUMBER: 60/016.892

; EARLIER FILING DATE: 1996-05-06

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9

; LENGTH: 26

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:4000 bp

; OTHER INFORMATION: (3631-7630) Primer 5

US-08-851-634-9

alignment_scores:

Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x US-08-851-634-9 ..

Align seg 1/1 to: US-08-851-634-9 from: 1 to: 26

55 LeuSerLysSerAlaThr 60

|||||

3 TTATCGAAATCAGCCACA 20

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-425-684-40

seq_documentation_block:

; Sequence 40, Application US/08425684

; Patent No. 5834252

; GENERAL INFORMATION:

; APPLICANT: STEMMER PH.D., WILLEM P.C.

; TITLE OF INVENTION: END-COMPLEMENTARY POLYMERASE REACTION

; NUMBER OF SEQUENCES: 136

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW

; STREET: ONE MARKET PLAZA, STEUART STREET TOWER, 20TH FLOOR

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: U.S.A.

; ZIP: 94105

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/425,684

; FILING DATE: 18-APR-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: DUNN ESQ., TRACY J.

; REGISTRATION NUMBER: 34,587

; REFERENCE/DOCKET NUMBER: 16528J-015400US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 326-2400

; TELEFAX: (415) 326-2422

; INFORMATION FOR SEQ ID NO: 40:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 40 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: circular

US-08-425-684-40

alignment_scores:

Quality: 6.00 Length: 6

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x US-08-425-684-40 ..

Align seg 1/1 to: US-08-425-684-40 from: 1 to: 40

45 LysValLeuLeuCysGly 50

|||||

5 AAAGTCTGCTATGTGC 22

seq_name: /cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-675-502-40

seq_documentation_block:

; Sequence 40, Application US/08675502

; Patent No. 5928905

US-09-332-522B-2_COPY_335_428 x 5223414-6 ..
Align seg 1/1 to: 5223414-6 from: 1 to: 41

55 LeuSerLysSerAlaThr 60
|||||
3 TTATCGAAATCAGCCACC 20

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-182-961B-3

seq_documentation_block:
; Sequence 3, Application US/08182961B
; Patent No. 5677125

GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.
; APPLICANT: PAGE, DAVID L.
; APPLICANT: OBERMILLER, PATRICE S.
; APPLICANT: ROBINSON-BENION, CHERYL L.
; TITLE OF INVENTION: METHOD OF DETECTION AND DIAGNOSIS OF PRE-INVASIVE CANC
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: I.C. WADDEY, JR.
; STREET: 27TH FLOOR, L & C TOWER, 401 CHURCH
; CITY: NASHVILLE
; STATE: TENNESSE
; COUNTRY: USA
; ZIP: 37219

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 kB storage
; COMPUTER: IBM PC/XT/AT compatible
; OPERATING SYSTEM: MS-DOS (version 5.0)
; SOFTWARE: WordPerfect 5.1/WordPerfect Editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/182,961B

FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: I.C. WADDEY, JR.
; REGISTRATION NUMBER: 25,180
; REFERENCE/DOCKET NUMBER: 0216-9409

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (615) 242-2400
; TELEFAX: (615) 242-2221
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens sapiens
; INDIVIDUAL ISOLATE: sample of non-comedo DCIS
; DEVELOPMENTAL STAGE: adult
; TISSUE TYPE: female breast
; CELL TYPE: ductal carcinoma in situ
; CELL LINE: not derived from a cell line
; ORGANELLE: no
; IMMEDIATE SOURCE:
; LIBRARY: cDNA library derived from human
; CLONE: obtained from identification of differential gene expression
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: unknown
; MAP POSITION: unknown
; UNITS: unknown
; FEATURE:

NAME/KEY: DCIS-3
; LOCATION: L27638
; IDENTIFICATION METHOD: microscopically-directed sampling and differential dis
; PUBLICATION INFORMATION: unpublished
; RELEVANT RESIDUES IN SEQ ID NO: 3
US-08-182-961B-3

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-182-961B-3/rev ..
Align seg 1/1 to reverse of: US-08-182-961B-3 from: 1 to: 46

55 LeuSerLysSerAlaThr 60
|||||
46 TTATCGAAATCAGCCACA 29

seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-864-473-50

seq_documentation_block:
; Sequence 50, Application US/08864473
; Patent No. 6027889

GENERAL INFORMATION:
; APPLICANT: Barany, Francis
; APPLICANT: Lubin, Matthew
; TITLE OF INVENTION: DETECTION OF NUCLEIC ACID SEQUENCE DIFFERENCES USING
; TITLE OF INVENTION: COUPLED LIGASE DETECTION AND POLYMERASE CHAIN REACTIONS
; FILE REFERENCE: 19603/441
; CURRENT APPLICATION NUMBER: US/08/864,473
; CURRENT FILING DATE: 1997-05-28
; EARLIER APPLICATION NUMBER: 60/018,532
; EARLIER FILING DATE: 1996-05-29
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 50
; LENGTH: 46
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Primer
; OTHER INFORMATION: Sequence
US-08-864-473-50

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-864-473-50/rev ..
Align seg 1/1 to reverse of: US-08-864-473-50 from: 1 to: 46

11 ThrThrIleIleValArg 16
|||||
30 ACCACGATCATAGTGGCG 13

seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-023-082A-45

seq_documentation_block:
; Sequence 45, Application US/09023082A
; Patent No. 6077692

GENERAL INFORMATION:
; APPLICANT: RUBEN, STEVEN M.
; APPLICANT: JIMENEZ, PABLO
; APPLICANT: DUAN, D. ROXANNE
; APPLICANT: RAMPI, MARK A.

APPLICANT: MENDRICK, DONNA
APPLICANT: ZHANG, JUN
APPLICANT: NI, JIAN
APPLICANT: MOORE, PAUL A.
APPLICANT: COLEMAN, TIMOTHY A.
APPLICANT: GRUBER, JOACHIM R.
APPLICANT: DILLON, PATRICK J.
APPLICANT: GENTZ, REINER L.
TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 09/023,082A
FILING DATE: 13-FEB-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/01790
FILING DATE: 14-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/461,195
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/023,852
FILING DATE: 13-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/039,045
FILING DATE: 28-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/862,432
FILING DATE: 23-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/910,875
FILING DATE: 13-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/055,561
FILING DATE: 13-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: STERNE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0360008/EKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-023-082A-45

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522b-2_COPY_335_428 x US-09-023-082A-45/rev ..
Align seg 1/1 to reverse of: US-09-023-082A-45 from: 1 to: 48

3 SerGlyGluThrMetSer 8
|||||
43 TCCGGGAACGATGCT 26

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-050-681A-3

seq_documentation_block:
; Sequence 3, Application US/08050681A
; Patent No. 5462854
; GENERAL INFORMATION:
; APPLICANT: Coassin, Peter J.
; APPLICANT: Cook, Donald M.
; APPLICANT: Rampal, Jang B.
; APPLICANT: Konrad, Kenneth D.
; TITLE OF INVENTION: Inverse Linkage
; TITLE OF INVENTION: Oligonucleotides for Chemical
; TITLE OF INVENTION: and Enzymatic Processes
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Beckman Instruments, Inc.
; STREET: 2500 Harbor Boulevard
; CITY: Fullerton
; STATE: California
; COUNTRY: USA
; ZIP: 92634
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb
; COMPUTER: IBM
; OPERATING SYSTEM: MS.DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,681A
; FILING DATE: 19930419
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Burgoon, Richard P.
; REGISTRATION NUMBER: 34,787
; REFERENCE/DOCKET NUMBER: 129D-1100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 773-7610
; TELEFAX: (714) 773-7936
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: no
; ANTI-SENSE: partial
; US-08-050-681A-3

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-050-681A-3 ..
Align seg 1/1 to: US-08-050-681A-3 from: 1 to: 50

55 LeuSerLysSerAlaThr 60
|||||
28 TATCGAATCAGCCACA 45

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-04275-3

seq_documentation_block:
; Sequence 3, Application PC/TUS9404275
; GENERAL INFORMATION:

```
; APPLICANT: Coassin, Peter J.
; APPLICANT: Cook, Donald M.
; APPLICANT: Rampal, Jang B.
; APPLICANT: Konrad, Kenneth D.
; TITLE OF INVENTION: Inverse Linkage
; TITLE OF INVENTION: Oligonucleotides for Chemical
; TITLE OF INVENTION: and Enzymatic Processes
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Beckman Instruments, Inc.
; STREET: 2500 Harbor Boulevard
; CITY: Fullerton
; STATE: California
; COUNTRY: USA
; ZIP: 92634
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb
; COMPUTER: IBM
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04275
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Burgoon, Richard P.
; REGISTRATION NUMBER: 34,787
; REFERENCE/DOCKET NUMBER: 129D-1100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 773-7610
; TELEFAX: (714) 773-7936
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: no
; ANTI-SENSE: partial
; PCT-US94-04275-3

alignment_scores:
  Quality: 6.00      Length: 6
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x PCT-US94-04275-3 ..
Align seg 1/1 to: PCT-US94-04275-3 from: 1 to: 50

55 LeuSerLysserAlaThr 60
|||||
28 TTATCGAAATCGGCACA 45

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-221-662-1

seq_documentation_block:
; Sequence 1, Application US/08221662
; Patent No. 5612199
; GENERAL INFORMATION:
; APPLICANT: Western, Linda M
; APPLICANT: Hahnenberger, Karen M
; APPLICANT: Rose, Samuel
; APPLICANT: Becker, Martin
; APPLICANT: Ullman, Edwin F
; TITLE OF INVENTION: METHOD FOR PRODUCING A POLYNUCLEOTIDE
; TITLE OF INVENTION: FOR USE IN SINGLE PRIMER AMPLIFICATION
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Syntex (U.S.A.) Inc.
; STREET: 3401 Hillview Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; STREET: 3401 Hillview Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/221,662
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/776,538
; FILING DATE: 11-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Leitner, Theodore J
; REGISTRATION NUMBER: 28,319
; REFERENCE/DOCKET NUMBER: 27410/ DO-1568
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 852-1091
; TELEFAX: (415) 496-3529
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-221-662-1

alignment_scores:
  Quality: 6.00      Length: 6
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-221-662-1 ..
Align seg 1/1 to: US-08-221-662-1 from: 1 to: 56

55 LeuSerLysserAlaThr 60
|||||
25 TTGTCGAAATCCGCACC 42

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-221-662-3

seq_documentation_block:
; Sequence 3, Application US/08221662
; Patent No. 5612199
; GENERAL INFORMATION:
; APPLICANT: Western, Linda M
; APPLICANT: Hahnenberger, Karen M
; APPLICANT: Rose, Samuel
; APPLICANT: Becker, Martin
; APPLICANT: Ullman, Edwin F
; TITLE OF INVENTION: METHOD FOR PRODUCING A POLYNUCLEOTIDE
; TITLE OF INVENTION: FOR USE IN SINGLE PRIMER AMPLIFICATION
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Syntex (U.S.A.) Inc.
; STREET: 3401 Hillview Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/221.662
; FILING DATE:
; CLASSIFICATION: 435
; APPLICATION NUMBER: 07/776.538
; FILING DATE: 11-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/776.538
; FILING DATE: 11-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Leitereg, Theodore J
; REGISTRATION NUMBER: 28,319
; REFERENCE/DOCKET NUMBER: 27410/ DO-1568
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 852-1091
; TELEFAX: (415) 496-3529
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 79 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-221-662-3

alignment_scores:
    Quality: 6.00      Length: 6
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-221-662-3 ..
Align seg 1/1 to: US-08-221-662-3 from: 1 to: 79
55 LeuSerLysserAlaThr 60
|||||
48 TTGTCGAAATCCGCGACC 65

seq_name: /cgn2_5/ptodata/2/ina/6_COMB.seq:us-09-023-082A-33

seq_documentation_block:
; Sequence 33, Application US/09023082A
; Patent No. 6077692
; GENERAL INFORMATION:
; APPLICANT: RUBEN, STEVEN M.
; APPLICANT: JIMENEZ, PABLO
; APPLICANT: DUAN, D. ROXANNE
; APPLICANT: RAMPY, MARK A.
; APPLICANT: MENDRICK, DONNA
; APPLICANT: ZHANG, JUN
; APPLICANT: NI, JIAN
; APPLICANT: MOORE, PAUL A.
; APPLICANT: COLEMAN, TIMOTHY A.
; APPLICANT: GRUBER, JOACHIM R.
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: GENTZ, REINER L.
; TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,082A
```

```
; FILING DATE: 13-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01790
; FILING DATE: 14-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/461,195
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/023,852
; FILING DATE: 13-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,045
; FILING DATE: 28-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,432
; FILING DATE: 23-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/910,875
; FILING DATE: 13-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,561
; FILING DATE: 13-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0360008/EKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-023-082A-33

alignment_scores:
    Quality: 6.00      Length: 6
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-09-023-082A-33/rev ..
Align seg 1/1 to reverse of: US-09-023-082A-33 from: 1 to: 80
3 SerGlyGluThrMetSer 8
|||||
24 TCCGGAGAAACCATGTCC 7

seq_name: /cgn2_5/ptodata/2/ina/3A_COMB.seq:us-08-182-961B-5

seq_documentation_block:
; Sequence 5, Application US/08182961B
; Patent No. 5677125
; GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.
; APPLICANT: PAGE, DAVID L.
; APPLICANT: OBERMILLER, PATRICE S.
; APPLICANT: ROBINSON-BENION, CHERYL L.
; TITLE OF INVENTION: METHOD OF DETECTION AND DIAGNOSIS OF PRE-INVASIVE CANC
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: I.C. WADDEY, JR.
; STREET: 27TH FLOOR, L & C TOWER, 401 CHURCH
; CITY: NASHVILLE
; STATE: TENNESSE
; COUNTRY: USA
```

ZIP: 37219
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
COMPUTER: IBM PC/XT/AT compatible
OPERATING SYSTEM: MS-DOS (version 5.0)
SOFTWARE: WordPerfect 5.1/WordPerfect Editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/182,961B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: I. C. WADDEY, JR.
REGISTRATION NUMBER: 25,180
REFERENCE/DOCKET NUMBER: 0216-9409
TELECOMMUNICATION INFORMATION:
TELEPHONE: (615) 242-2400
TELEFAX: (615) 242-2221
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 84
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens sapiens
INDIVIDUAL ISOLATE: sample of non-comedo DCIS
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: ductal carcinoma in situ
CELL LINE: not derived from a cell line
ORGANELLE: no
IMMEDIATE SOURCE:
LIBRARY: cDNA library derived from human
CLONE: obtained from identification of differential gene expression
POSITION IN GENOME:
CHROMOSOME/SEGMENT: unknown
MAP POSITION: unknown
UNITS: unknown
FEATURE:
NAME/KEY: DCIS-5
LOCATION: L27641
IDENTIFICATION METHOD: microscopically-directed sampling and differential display
PUBLICATION INFORMATION: unpublished
RELEVANT RESIDUES IN SEQ ID NO: 5
US-08-182-961B-5

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-182-961B-5 ..
Align seg 1/1 to: US-08-182-961B-5 from: 1 to: 84

55 LeuserLysSerAlaThr 60
|||||
9 TTTATCGAATACGCCACA 26

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-009-265-38
seq_documentation_block:
Sequence 38, Application US/08009265
Patent No. 5547871

GENERAL INFORMATION:
APPLICANT: Black Dr., Bruce C.
TITLE OF INVENTION: Heterologous Signal Sequences For
Secretion Of Insect Controlling Proteins
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: 1937 West Main Street, P.O. Box 60
CITY: Stamford
STATE: CT
COUNTRY: USA
ZIP: 06904-0060
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/009,265
FILING DATE: 19930125
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gordon Mr., Alan M.
REGISTRATION NUMBER: 30637
REFERENCE/DOCKET NUMBER: 31868-00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 203-321-2719
TELEFAX: 203-321-2971
TELEX: 710-474-4059
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bombyx mori
US-08-009-265-38

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-009-265-38 ..
Align seg 1/1 to: US-08-009-265-38 from: 1 to: 90

20 ThrGluArgArgThrAla 25
|||||
71 ACGGAAGAAGAACGGCT 88

seq_name: /cgn2_6/ptodata/2/1na/PCTUS_COMB.seq:PCT-US94-06456-18

seq_documentation_block:
Sequence 18, Application PC/TUS9406456
GENERAL INFORMATION:

APPLICANT: Beutel, Bruce A.
APPLICANT: Coppola, George R.
APPLICANT: Sherman, Michael I.
TITLE OF INVENTION: Oligonucleotides Which Inhibit HIV Protease Function
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: New Jersey

```
;
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC - DOS
; SOFTWARE: DW4.V2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/06456
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/073,873
; FILING DATE: 09-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliott M.
; REGISTRATION NUMBER: 24,025
; REFERENCE/DOCKET NUMBER: 23550-89
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 98 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: oligonucleotide
; PCT-US94-06456-18

alignment_scores:
    Quality: 6.00      Length: 6
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x PCT-US94-06456-18 ..
Align seg 1/1 to: PCT-US94-06456-18 from: 1 to: 98

12 ThrilleValArgArg 17
|||||
57 ACGAUUAUAGUAGACGC 74

seq_name: /cgn2.6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-06456-18

seq_documentation_block:
; Sequence 47, Application PC/TUS9406456
; GENERAL INFORMATION:
; APPLICANT: Beutel, Bruce A.
; APPLICANT: Coppola, George R.
; APPLICANT: Sherman, Michael I.
; TITLE OF INVENTION: Oligonucleotides Which Inhibit HIV Protease Function
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC - DOS
; SOFTWARE: DW4.V2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/06456
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/073,873

;
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC - DOS
; SOFTWARE: DW4.V2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/06456
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/073,873

;
; FILING DATE: 09-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliott M.
; REGISTRATION NUMBER: 24,025
; REFERENCE/DOCKET NUMBER: 23550-89
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 98 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: oligonucleotide
; PCT-US94-06456-47

alignment_scores:
    Quality: 6.00      Length: 6
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x PCT-US94-06456-47 ..
Align seg 1/1 to: PCT-US94-06456-47 from: 1 to: 98

12 ThrilleValArgArg 17
|||||
57 ACGATTATAGTAAGACGC 74

seq_name: /cgn2.6/ptodata/2/ina/5A_COMB.seq:US-07-648-796A-18

seq_documentation_block:
; Sequence 18, Application US/07648796A
; Patent No. 5310876
; GENERAL INFORMATION:
; APPLICANT: Dr. Hubert Bayer
; APPLICANT: Dr. Erhard Kopetzki
; TITLE OF INVENTION: Expression of HIV1 and HIV2
; TITLE OF INVENTION: Polypeptides and their use
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felie & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/648,796A
; FILING DATE: 19910125
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P 40 02 636.1
; FILING DATE: 30-JANUARY-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, NG. 5310876man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: HUBR 784
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 base pairs
; TYPE: NUCLEIC ACID
```



```
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-648-796A-18

alignment_scores:
  Quality: 6.00 Length: 6
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-07-648-796A-18/rev ..
Align seg 1/1 to reverse of: US-07-648-796A-18 from: 1 to: 100

55 LeuSerLysSerAlaThr 60
76 CTGCTAAAGCGCAACC 59

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5223424-14

seq_documentation_block:
; Patent No. 5223424
; APPLICANT: COCHRAN, MARK; CHIANG, CHRISTINA H.; MACDONALD,
; RICHARD D.
; TITLE OF INVENTION: ATTENUATED HERPESVIRUSES AND
; HERPESVIRUSES WHICH INCLUDE FOREIGN DNA ENCODING AN AMINO
; ACID SEQUENCE
; NUMBER OF SEQUENCES: 16
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/225,032
; FILING DATE: 27-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 78,519
; FILING DATE: 27-JUL-1987
; APPLICATION NUMBER: 933,107
; FILING DATE: 20-NOV-1986
; APPLICATION NUMBER: 902,887
; FILING DATE: 02-SEP-1986
; APPLICATION NUMBER: 887,140
; FILING DATE: 17-JUL-1986
; APPLICATION NUMBER: 823,102
; FILING DATE: 27-JAN-1986
; APPLICATION NUMBER: 773,430
; FILING DATE: 06-SEP-1985
; SEQ ID NO:14:
; LENGTH: 114
5223424-14

alignment_scores:
  Quality: 6.00 Length: 6
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x 5223424-14/rev ..
Align seg 1/1 to reverse of: 5223424-14 from: 1 to: 114

32 LysTyrArgCysSerIle 37
87 AATACAGGTGCAGTATT 70

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-751-304B-2

seq_documentation_block:
; Sequence 2, Application US/07751304B
; Patent No. 5424411
; GENERAL INFORMATION:
; APPLICANT: YAGI, Shintaro
; APPLICANT: TANAKA, Kiyoko
; APPLICANT: YOSHIOKA, Juri
```

```
; APPLICANT: SUZUKI, Masanori
; TITLE OF INVENTION: E.COLI-DERIVED UPSTREAM REGULATORY
; TITLE OF INVENTION: SEQUENCE OPERABLE IN YEAST
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEGNER, CANTOR, MUELLER & PLAYER
; STREET: 1233 20th Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-8218
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/751,304B
; FILING DATE: 28-AUG-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 226566/1992
; FILING DATE: 30-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueller, Douglas P.
; REGISTRATION NUMBER: 30,300
; REFERENCE/DOCKET NUMBER: P-450-23021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-0400
; TELEFAX: (202) 835-0605
; TELEX: 440706 and 248394
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 153 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-751-304B-2

alignment_scores:
  Quality: 6.00 Length: 6
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-07-751-304B-2/rev ..
Align seg 1/1 to reverse of: US-07-751-304B-2 from: 1 to: 153

56 SerLysSerAlaThrLeu 61
|||||
50 TCCAAATCTGCAACGCTG 33

seq_name: /cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-633-879C-7

seq_documentation_block:
; Sequence 7, Application US/08633879C
; Patent No. 5928922
; GENERAL INFORMATION:
; APPLICANT: Kivirikko, Kari I.
; APPLICANT: Pihlajaniemi, Taina
; APPLICANT: Helaakoski, Tarja I.
; APPLICANT: Annunen, Pia P.
; APPLICANT: Nissi, Ritva K.
; APPLICANT: No. 5928922elainen, Minna K.
; TITLE OF INVENTION: 2 SUBUNIT OF PROLYL-4-HYDROXYLASE
; NUCLEIC ACID SEQUENCES ENCODING SUCH SUBUNIT AND
; TITLE OF INVENTION: METHODS FOR PRODUCING THE SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
```

```
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,879C
; FILING DATE: 10-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,879C
; FILING DATE: 10-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0041-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 3...95
; OTHER INFORMATION:
US-08-633-879C-7

alignment_scores:
    Quality: 6.00      Length: 6
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-633-879C-7 ..
Align seg 1/1 to: US-08-633-879C-7 from: 1 to: 200

52 GluAlaLysLeuSerLys 57
|||||
72 GAAGCCCAAGCTTCCCAAG 89

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-193-078B-19

seq_documentation_block:
; Sequence 19, Application US/08193078B
; Patent No. 5846757
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark
; APPLICANT: Feldman, Daniel
; APPLICANT: McCue, Ann
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWN, MARTIN, HALLER & McCLAIN
; STREET: 1660 UNION STREET
; CITY: SAN DIEGO
; STATE: CA
; COUNTRY: USA

; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 92101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/193,078B
; FILING DATE: 07-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/868,354
; FILING DATE: 10-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/745,206
; FILING DATE: 15-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-53607
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-193-078B-19

alignment_scores:
    Quality: 6.00      Length: 6
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-193-078B-19 ..
Align seg 1/1 to: US-08-193-078B-19 from: 1 to: 249

7 MetSerGinglyThrThr 12
|||||
184 ATGAGTCAGGGAACAACG 201

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-171-385-9

seq_documentation_block:
; Sequence 9, Application US/08171385
; Patent No. 5527884
; GENERAL INFORMATION:
; APPLICANT: Mary E. Russell
; APPLICANT: Ulrike Utans
; TITLE OF INVENTION: Mediators of Chronic Allograft
; REJECTION
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/171,385
; FILING DATE:
```

```
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/006001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 274
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-171-385-9

alignment_scores:
    Quality: 6.00      Length: 6
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-171-385-9/rev ..
Align seg 1/1 to reverse of: US-08-171-385-9 from: 1 to: 274

77 ValLeuLyShisHisVal 82
|||||
243 GTTCTCAAGCACCATTGTT 226

seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-361-441B-9

seq_documentation_block:
; Sequence 9, Application US/08361441B
; Patent No. 6077948
; GENERAL INFORMATION:
; APPLICANT: Russell, Mary E.
; APPLICANT: Utans, Ulrike
; TITLE OF INVENTION: MEDIATORS OF CHRONIC ALLOGRAFT REJECTION
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,441B
; FILING DATE: 21-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/171,385
; FILING DATE: 21-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/014001
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 274 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-361-441B-9

alignment_scores:
    Quality: 6.00      Length: 6
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-361-441B-9/rev ..
Align seg 1/1 to reverse of: US-08-361-441B-9 from: 1 to: 274

77 ValLeuLyShisHisVal 82
|||||
243 GTTCTCAAGCACCATTGTT 226

seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-994-946A-4

seq_documentation_block:
; Sequence 4, Application US/08994946A
; Patent No. 6046317
; GENERAL INFORMATION:
; APPLICANT: Koulu, Markku
; APPLICANT: Karvonen, Matti
; APPLICANT: Pesonen, Ullamari
; APPLICANT: Uusitupa, Matti
; TITLE OF INVENTION: A DNA Molecule Encoding a Mutant
; TITLE OF INVENTION: Prepro-Neuropeptide Y, a Mutant Signal Peptide, and Uses
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 13th Street NW, Suite 701-E
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/994,946A
; FILING DATE: 19-DEC-1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 2328-110
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-783-6040
; TELEFAX: 202-783-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 300 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-994-946A-4

alignment_scores:
    Quality: 6.00      Length: 6
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000
```

```

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-994-946A-4/rev ..
Align seg 1/1 to reverse of: US-08-994-946A-4 from: 1 to: 300

86 ArgLysThrLeuGlnAsn 91
|||||
180 AGAAGAACATTCAGAAAT 163

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-171-385-11

seq_documentation_block:
; Sequence 11, Application US/08171385
; Patent No. 5527884
; GENERAL INFORMATION:
; APPLICANT: Mary E. Russell
; APPLICANT: Ulrike Utans
; TITLE OF INVENTION: Mediators of Chronic Allograft
; TITLE OF INVENTION: Rejection
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/171.385
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/006001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-171-385-11

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-171-385-11/rev ..
Align seg 1/1 to reverse of: US-08-171-385-11 from: 1 to: 309

77 ValLeuLysHisHisVal 82
|||||
253 GTTCACGACCATGTT 236

seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-361-441B-11

seq_documentation_block:
; Sequence 11, Application US/08171385
; Patent No. 5527884
; GENERAL INFORMATION:
; APPLICANT: Mary E. Russell
; APPLICANT: Ulrike Utans
; TITLE OF INVENTION: Mediators of Chronic Allograft
; TITLE OF INVENTION: Rejection
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/361.441B
; FILING DATE: 21-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/171,385
; FILING DATE: 21-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/014001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-361-441B-11

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-361-441B-11/rev ..
Align seg 1/1 to reverse of: US-08-361-441B-11 from: 1 to: 309

77 ValLeuLysHisHisVal 82
|||||
253 GTTCACGACCATGTT 236

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-171-385-18

seq_documentation_block:
; Sequence 18, Application US/08171385
; Patent No. 5527884
; GENERAL INFORMATION:
; APPLICANT: Mary E. Russell
; APPLICANT: Ulrike Utans
; TITLE OF INVENTION: Mediators of Chronic Allograft
; TITLE OF INVENTION: Rejection
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.

```

```

; Sequence 11, Application US/08361441B
; Patent No. 6077948
; GENERAL INFORMATION:
; APPLICANT: Russell, Mary E.
; APPLICANT: Utans, Ulrike
; TITLE OF INVENTION: MEDIATORS OF CHRONIC ALLOGRAFT REJECTION
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,441B
; FILING DATE: 21-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/171,385
; FILING DATE: 21-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/014001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-361-441B-11

alignment_scores:
Quality: 6.00 Length: 6
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x US-08-361-441B-11/rev ..
Align seg 1/1 to reverse of: US-08-361-441B-11 from: 1 to: 309

77 ValLeuLysHisHisVal 82
|||||
253 GTTCACGACCATGTT 236

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-171-385-18

seq_documentation_block:
; Sequence 18, Application US/08171385
; Patent No. 5527884
; GENERAL INFORMATION:
; APPLICANT: Mary E. Russell
; APPLICANT: Ulrike Utans
; TITLE OF INVENTION: Mediators of Chronic Allograft
; TITLE OF INVENTION: Rejection
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.

```

```
;
;
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/171,385
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/006001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 320
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-171-385-18

alignment_scores:
  Quality: 6.00 Length: 6
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
  US-09-332-522B-2_COPY_335_428 x US-08-361-441B-18/rev ..
  Align seg 1/1 to reverse of: US-08-361-441B-18 from: 1 to: 320
  77 ValLeuLySHisHisVal 82
  |||||
  250 GTTCTCAAGCACCATGTT 233

seq_name: /cgn2_6/ptodata/2/1na/6_COMB.seq:US-08-361-441B-18

seq_documentation_block:
; Sequence 18, Application US/08361441B
; Patent No. 6077948
; GENERAL INFORMATION:
; APPLICANT: Russeil, Mary E.
; APPLICANT: Utans, Ulrike
; TITLE OF INVENTION: MEDIATORS OF CHRONIC ALLOGRAFT REJECTION
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,441B
; FILING DATE: 21-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/171,385
; FILING DATE: 21-DEC-1993
; ATTORNEY/AGENT INFORMATION:
```


OM of: US-09-332-522B-2_COPY_335_428 to: EST:* out_format : pfs

Date: Sep 27, 2000 1:46 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEJ-frame+ p2n.model -DEV=xl
-Q/cn2_1/USPTO.spool/US09332522/runat_27092000_091221_6524/app_query.fasta_1.1355
-DB=EST -CFMT=fastp -SUFFIX=oli.rst -GAPOP=4.500 -GAPEXT=0.050
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-CGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000 -DELOP=6.000
-DELEXT=7.000 -XGAPOP=60.000 -XGAPEXT=60.000 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=45
-MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09332522 -CGN1_1_2977 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-332-522B-2_COPY_335_428

Query length: 94

Database: EST:*

Database sequences: 5247842

Database length: -2090053206

Search time (sec): 1353.190000

WARN: XGAPOP and XGAPEXT must be equal. Assuming XGAPOP=XGAPEXT=60.000
WARN: XGAPEXT and XGAPEXT must be equal. Assuming XGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_est47:D35004	+	35.00	708.78	360	D35004 CELK015H2F Yuji Kohara u
gb_g988:AQ552484	-	9.00	169.16	451	A0552484 RPCI-11-434H18 TV RPCI
gb_est34:AV283137	-	8.00	154.45	183	AV283137 AV283137 RIKEN full-1e
gb_est46:AW02646	-	8.00	151.28	295	AW02646 TgESTz287b06.y1 TGVEG
gb_g982:AQ085039	-	8.00	149.57	382	AQ085039 HS_2265_A2-D08 MF CIT
gb_est9:AA19166	-	8.00	148.64	439	AA19166 v073b02.r1 Barstead m
gb_est40:AW204688	-	8.00	148.51	448	AW204688 UI-H-BW0-ail-h-06-0-UI
gb_est46:AW072855	-	8.00	148.46	451	AW072855 TgESTz289f03.y1 TGVEG
gb_g983:AQ237040	+	8.00	146.76	24.03	AQ237040 RPCI11-71J1 TK RPCI-11
gb_g981:AQ753213	+	8.00	146.76	24.03	AQ753213 HS_5308_A2-H06 T7A RPC
gb_g9815:AQ073363	+	8.00	145.81	27.13	AQ073363 RPCI-23-393022.TJB RPC
gb_g9815:AQ048465	-	8.00	140.93	50.72	A2048465 PSB141 Barley PstI 9e
gb_g9813:AA891352	-	7.00	136.23	92.66	AA891352 EST195155 Normalized r
gb_est8:AA474206	+	7.00	135.17	106.28	AA474206 ve54g10.r1 Beddington
gb_est31:AV164024	+	7.00	133.27	135.60	AV164024 AV164024 Mus musculus
gb_est28:AV282728	-	7.00	133.20	136.77	AV282728 AA3A-ABB-D-05-R Anophe
gb_est46:AL100958	-	7.00	132.97	140.86	AL100958 EST210247 Normalized r
gb_est30:AV067361	-	7.00	132.97	140.86	AV067361 AV067361 Mus musculus
gb_est23:AI676373	-	7.00	132.81	143.77	AI676373 etmEST0124 Eth1 Elmer1
gb_est34:AV293889	+	7.00	132.41	151.28	AV293889 AV293889 RIKEN full-1e
gb_est3:AA199336	+	7.00	132.04	158.73	AA199336 mt55f04.r1 Stratagene
gb_est45:AW641622	+	7.00	131.95	160.44	AW641622 cm09d12.w1 Blackshear
gb_est46:AA897499	+	7.00	131.79	163.85	AA897499 EST266942 tomato ovary
gb_g9818:FR0013555	+	7.00	131.74	164.98	AL004804 F.rubripes GSS sequenc
gb_est33:AV274480	+	7.00	131.71	165.55	AV274480 AV274480 RIKEN full-1e
gb_est53:W14668	+	7.00	131.45	171.19	W14668 mb30f02.r1 Soares mouse
gb_est33:AV269151	+	7.00	131.40	172.32	AV269151 AV269151 RIKEN full-1e
gb_est41:AW332840	+	7.00	131.30	174.57	AW332840 SL3H1 AGS-1 Pneumocyst
gb_g9814:AC002831	+	7.00	131.17	177.37	A202831 RPCI-23-344L19.TJ RPCI
gb_est46:C17181	-	7.00	131.12	178.49	C17181 C17181 Clontech human ad
gb_est49:H65208	+	7.00	131.12	178.49	H65208 yu66f02.s1 Weizmann Olf
gb_est40:AW298944	+	7.00	130.88	184.06	AW298944 262 MARC 1BOV Bos taur
gb_est55:242367	+	7.00	130.68	189.05	242367 HSC06G121 normalized inf
gb_est16:AA106092	+	7.00	130.56	191.82	AA106092 cb02f06.t3 Zf adult he
gb_est12:AA786770	+	7.00	130.47	194.02	AA786770 msh12al.r1 Aspergillus
gb_est46:C30348	-	7.00	130.47	194.02	C30348 C30348 Yuji Kohara unpub
gb_est47:C53448	-	7.00	130.47	194.02	C53448 C53448 Yuji Kohara unpub
gb_est34:AV298105	-	7.00	130.43	195.13	AV298105 AV298105 RIKEN full-1e
gb_est8:AA480552	+	7.00	130.36	196.78	AA480552 nc70b002.r1 NCI_CGAP_P

gb_est31:AV145949 + 7.00 130.25 199.52 310 ! AV145949 AV145949 Mus muscul
gb_est5:AA356000 + 7.00 130.21 200.62 312 ! AA356000 EST84653 Jurkat T-c
gb_est13:AA900350 - 7.00 130.17 201.71 314 ! AA900350 UI-R-E0-cp-h-01-0-U
gb_est53:W06635 + 7.00 130.13 202.80 316 ! W06635 T2197 MVAR4 bloodstre
gb_est49:H64717 - 7.00 130.11 203.35 317 ! H64717 yu66f02.r1 Weizmann O
gb_est54:W85333 - 7.00 130.00 206.08 322 ! W85333 mf46d10.r1 Soares mou

seq_name: gb_est47:D35004

seq_documentation_block:
LOCUS D35004 360 bp mRNA EST 08-AUG-1994
DEFINITION CELK015H2F Yuji Kohara unpublished cDNA Caenorhabditis elegans CDNA
clone yk15h12 5', mRNA sequence.
ACCESSION D35004
VERSION D35004.1 GI:526520
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
Tabara,H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT On Apr 7, 1998 this sequence version replaced gi:3036136.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
source 1..360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(el489)"
/db_xref="taxon:6239"
/clone="yk15h12"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT 123 a 63 c 100 g 73 t 1 others
ORIGIN

alignment_scores:
Quality: 35.00 Length: 35
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x D35004

Align seg 1/1 to: D35004 from: 1 to: 360

1 SerAspSerGlyGluThrMetSerGlnGlyThrThrIleIleValArgAr 17
|||||
254 AGTGATTCGGGGAGACTATGTCACGGAGACTACTATATGTCGAAG 303

17 gProLysThrGluArgArgThrAlaHisAsnLeuIleGlyLysIleTyrA 34
|||||
304 ACCAAAACCGAGCGTCGTACGGGACACACATCTCATCGAAAAGAGTATA 353

34 rgCys 35
|||||
354 GATGC 358

seq_name: gb_gss8:AQ552484
seq_documentation_block:
LOCUS AQ552484 451 bp DNA GSS 28-MAY-1999

```

DEFINITION   RPCI-11-434H18_TV RPCI-11 Homo sapiens genomic clone RPCI-11-
ACCESSION    434H18, genomic survey sequence.
VERSION      AQ552484.1 GI:4911661
KEYWORDS     GSS.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 451)
AUTHORS      Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
              Venter,J.C.
TITLE        Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
              Map Building
JOURNAL      Unpublished (1997)
COMMENT      On Dec 15, 1999 this sequence version replaced gi:4211931.
              Other GSSs: RPCI-11-434H18.TJ
              Contact: Shaying Zhao, William Nierman, Mark Adams
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: hbe@tigr.org
              Clones are derived from the human BAC library RPCI-11. For BAC
              library availability, please contact Pieter de Jong
              (pieter@dejong.med.buffalo.edu). Clones may be purchased from
              BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
              Research Genet cs (info@resgen.com). BAC end search page:
              http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
              Seq primer: P7
              Class: BAC ends.
FEATURES     Location/Qualifiers
             1..451
             /organism="Homo sapiens"
             /db_xref="GDB:7666457"
             /db_xref="taxon:9606"
             /clone="RPCI-11-434H18"
             /clone_lib="RPCI-11"
             /sex="Male"
             /cell_type="Lymphocytes"
             /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
              RPCI11 Human Male BAC Library"
BASE COUNT   74 a 83 c 77 g 217 t
ORIGIN
alignment_scores:
  Quality: 9.00 Length: 9
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-332-522B-2_COPY_335_428 x AQ552484/rev ...
Align seg 1/1 to reverse of: AQ552484 from: 1 to: 451
85 MetArgLysThrLeuGlnAsnArg 93
|||||
281 ATGAGAAAACCTTCAAAACATCGA 255
seq_name: gb_est34:AV283137
seq_documentation_block:
LOCUS      AV283137 183 bp mRNA EST
DEFINITION AV283137 RIKEN full-length enriched, adult male testis (DH10B) Mus
            musculus cDNA clone 4933433102 3' similar to AF088868 Homo sapiens
            fibrousheathln II mRNA, mRNA sequence.
ACCESSION  AV283137
VERSION     AV283137.1 GI:6275886
KEYWORDS    EST.
SOURCE      house mouse.

```

```

ORGANISM     Mus musculus
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
              1 (bases 1 to 183)
AUTHORS      Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
              Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
              Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
              Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
              Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
              Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
              Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
              Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tomioka,N.,
              Tsunoda,T., Watanabe,S., Yamamura,T., Yasunishi,A.,
              Yokota,T., Yoshiki,A., Yoshino,Y., Muramatsu,M. and Hayashizaki,Y.
              RIKEN Mouse ESTs (Konno,H., et al.)
              Unpublished (1999)
              On Apr 7, 1998 this sequence version replaced gi:3036263.
              Contact: Yoshihide Hayashizaki
              Genome Exploration Research Group, Life Science Tsukuba Center,
              Genome Science Laboratory
              The Institute of Physical and Chemical Research (RIKEN), Genomic
              Sciences Center
              3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
              Tel: +81-298-36-9013
              Fax: +81-298-36-9098
              Email: genome-res@rtc.riken.go.jp,
              URL: http://genome.rtc.riken.go.jp/
              Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
              Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
              Hayashizaki,Y.
              Transcriptional sequencing: A method for DNA sequencing using RNA
              polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
              Itoh,M., Kitsumai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
              Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
              Okazaki,Y. and Hayashizaki,Y.
              Automated filtration-based high-throughput plasmid preparation
              system. Genome Res. 9 (5), 463-470 (1999)
              Carninci,P. and Hayashizaki,Y.
              High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
              19-44 (1999)
              Please visit our web site (http://genome.rtc.riken.go.jp) for
              further details.
              Location/Qualifiers
              1..183
              /organism="Mus musculus"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone="4933433102"
              /clone_lib="RIKEN full-length enriched, adult male testis
              (DH10B)"
              /sex="male"
              /tissue_type="testis"
              /dev_stage="adult"
              /lab_host="DH10B"
              /note="Site_1: Sall; Site_2: BamHI; cDNA library was
              prepared and sequenced in Mouse Genome Encyclopedia
              Project of Genome Exploration Research Group in Riken
              Genomic Sciences Center and Genome Science Laboratory in
              RIKEN. Division of Experimental Animal Research in Riken
              contributed to prepare mouse tissues. 1st strand cDNA was
              primed with a primer [5'-
              GAGAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was
              prepared by using trihalose thermo-activated reverse
              transcriptase and subsequently enriched for full-length by
              cap-trapper. Second strand cDNA was prepared with the
              primer adapter of sequence [5'-
              GAGAGAGAGATCCAGATTAATAATATCCCTCCCCCCCC 3']. cDNA
              was cloned into the XhoI and BamHI sites. Vector: a
              modified pBluescript KS(+) after bulk excision from Lambda
              FLC I. Cloning sites, 5' end: Sall; 3' end: BamHI."
BASE COUNT   51 a 41 c 38 g 53 t
ORIGIN

```


alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x AV283137/rev ..

Align seg 1/1 to reverse of: AV283137 from: 1 to: 183

44 LeuLysValLeuLeuCysGlyAsp 51
|||||
103 TTAAGGTTCTGCTCGGGAGAT 80

seq_name: gb_est46:AW702646

seq_documentation_block:
LOCUS AW702646 295 bp mRNA EST 18-APR-2000
DEFINITION TgESTz87b06.y1 TgVEG-Tachyzoite cDNA Toxoplasma gondii cDNA clone
TgESTz87b06.y1 5', mRNA sequence.

ACCESSION AW702646
VERSION AW702646.1 GI:7586808
KEYWORDS EST.
SOURCE Toxoplasma gondii.
ORGANISM Toxoplasma gondii.
Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeriida;
Sarcocystidae; Toxoplasma.
REFERENCE 1 (bases 1 to 295)
AUTHORS Hehl,A., Manger,I., Marra,M., Sibley,L.D., Ajioke,J.A.,
Aslett,M.A., Dietrich,N., Dubucque,T., Hillier,L., Kucaba,T.,
Wan,K.L., Waterston,R.H. and Boothroyd,J.
WashU-Merck-Stanford-NIH Toxoplasma EST project
Unpublished (1996)
On Nov 2, 1998 this sequence version replaced gi:3830448.
Contact: Marra M
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@watson.wustl.edu
Contact David Sibley (toxost@borcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco
High quality sequence stop: 291.

FEATURES
Location/Qualifiers
1..295
/organism="Toxoplasma gondii"
/strain="VEG (Type III)"
/db_xref="taxon:5811"
/clone="TgESTz87b06.y1"
/clone_lib="TgVEG-Tachyzoite cDNA"
/dev_stage="Tachyzoite"
/lab_host="SOLR cells"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was synthesized from poly mRNA using an
oligo-dt primer containing an xhoI site. Following second
strand synthesis, EcoRI adapters were ligated to the cDNA
and products were size- selected on sephacryl S500. The
cDNAs were ligated to EcoRI/ XhoI prepared lambda ZapII
(Stratagene). Clones were converted to phagemids by mass
excision using ExAssist helper phage and E.coli SOLR cells
(Stratagene). Insert sizes range from 0.3-3.0 kb. The
library may contain a small percentage of host or
bacterial contaminants. NOTE: Many clones lack the 5'
EcoRI site but contain the upstream polylinker sites."

BASE COUNT 77 a 78 c 100 g 38 t 2 others

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x AW702646

Align seg 1/1 to: 295

17 ArgProLysThrGluArgArgThr 24
|||||
179 AGCGCAAAACGGAAGCGAACG 202

seq_name: gb_gss2:AQ085039

seq_documentation_block:
LOCUS AQ085039 382 bp DNA GSS 26-AUG-1998
DEFINITION HS_2265_A2_D08_MF CIT Approved Human Genomic Sperm Library D
Homo sapiens genomic clone Plate-2265 Col-16 Row-G, genomic
survey sequence.

ACCESSION AQ085039
VERSION AQ085039.1 GI:3454256
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 382)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2265 row: G column: 16
Class: BAC ends
High quality sequence stop: 382.

FEATURES
Location/Qualifiers
1..382
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-2265 Col-16 Row-G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT 173 a 48 c 58 g 100 t 3 others

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x AQ085039/rev ..

Align seg 1/1 to reverse of: AQ085039 from: 1 to: 382

54 LysLeuSerLysSerAlaThrLeu 61
|||||
97 AAAGTGCCTCAATCTGCACCTTG 74

```

seq_name: gb_est9:AA619166

seq_documentation_block:
LOCUS      439 bp      mRNA      EST      09-OCT-1997
DEFINITION      vo73b02.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
IMAGE:1064715 5', mRNA sequence.
ACCESSION      AA619166
VERSION        AA619166
KEYWORDS       EST.
SOURCE         AA619166.1 GI:2523042
              house mouse.
ORGANISM       Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE      1 (bases 1 to 439)
AUTHORS        Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
              Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
              Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
              Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
              Waterston,R.
TITLE          The WashU-HMI Mouse EST Project
JOURNAL        Unpublished (1996)
COMMENT        Contact: Marra M/Mouse EST Project
              WashU-HMI Mouse EST Project
              Washington University School of MedicineP
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: mouseest@watson.wustl.edu
              This clone is available royalty-free through LNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              MGI:587075
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 414.
Location/Qualifiers
1. 439
/organism="Mus musculus"
/strain="C3H"
/db_xref="taxon:10090"
/clone="IMAGE:1064715"
/clone_lib="Barstead mouse myotubes MPLRB5"
/cell_lines="C2C12"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker. Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCCCTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGATCCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead. The C2C12 cell line
(available from ARCC, catalog # CRU-1772) differentiates
rapidly, forming contractile myotubes and producing
characteristic muscle proteins."
BASE COUNT      103 a      68 c      100 g      168 t
ORIGIN

alignment_scores:
Quality:      8.00      Length:
Ratio:      1.000      Gaps:
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
us-09-332-522b-2_copy_335_428 x AA619166/rev ..
Align seg 1/1 to reverse of: AA619166 from: 1 to: 439
10 GlyThrThrIleIleValArgarg 17
|||||
366 GGCAACAACATCATAGTCGACGAG 343

seq_name: gb_est40:AW294688

seq_documentation_block:
LOCUS      448 bp      mRNA      EST      16-JAN-2000
DEFINITION      UI-H-BWO-all-h-06-0-UI.s1 NCI_CGAP_Sub6 Homo sapiens cDNA clone
IMAGE:2729723 3', mRNA sequence.
ACCESSION      AW294688
VERSION        AW294688.1 GI:6701324
KEYWORDS       EST.
SOURCE         AW294688.1
              human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 448)
AUTHORS        NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL        Unpublished (1997)
COMMENT        On Aug 21, 1998 this sequence version replaced gi:3704708.
              Contact: Robert Strausberg, Ph.D.
              Tel: (301) 496-1550
              Email: Robert.Strausberg@nih.gov
              The sequence contained an oligo-dT track that was present in the
              oligonucleotide that was used to prime the synthesis of first
              strand cDNA and therefore this may represent a bonafide poly A
              tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
              NCI-CGAP clone distribution information can be found through the
              I.M.A.G.E. Consortium/LLNL at:
              www-bio.llnl.gov/bbrp/image/image.html
              Seq primer: M13 Forward
              POLYA=Yes.
Location/Qualifiers
1. 448
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2729723"
/clone_lib="NCI_CGAP_Sub6"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub6
is a subtracted library derived from BW, which consists of
a mixture of four normalized libraries: NCI_CGAP_Brn50,
NCI_CGAP_Lu13, NCI_CGAP_Ov18, GBC1. The NCI_CGAP_Sub6
library had 7 million recombinants. A single-stranded DNA
preparation of BW was used as a tracer in a subtractive
hybridization with a driver comprising: the IMAGE pool
(NCI_CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,
1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE ClonesIDs
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI_CGAP_Lu5 pool 1 LLAM 3573-3582, 3851-3854 (IMAGE
ClonesIDs 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
ClonesIDs 1257096-1258631, 1469064-1470983,
1475592-1476743); NCI_CGAP_Pr22 pool 1 LLAM 2457-2459,
2758-2759, 3062-3068 (IMAGE ClonesIDs 985608-986759,
1101192-1101959, 1217928-1220615); NCI_CGAP_Co10 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE ClonesIDs
1057416-1061255, 1144584-1145351). (50% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE ClonesIDs 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE ClonesIDs 2710536-2714555) (20% of
the driver population), plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE ClonesIDs 2712456-2723591) (30% of
the driver population). Subtraction was performed as
previously described [Bonaldo, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806.;
TAG_LIB=GBC1; TAG_TRISUE=B cells germinal; TAG_SEQ=TCA"
BASE COUNT      148 a      87 c      79 g      133 t
ORIGIN

alignment_scores:

```

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x AW294688/rev ..
Align seg 1/1 to reverse of: AW294688 from: 1 to: 448

41 IlegInLnLeuLysValleuLeu 48
|||||
408 ATACAACAATTGAAAGTGTGCTC 385

seq_name: gb_est46:AW702855

seq_documentation_block:
LOCUS AW702855 451 bp mRNA EST 18-APR-2000
DEFINITION TgESTz289f03.y1 TgVEG-Tachyzoite cDNA Toxoplasma gondii cDNA clone
ACCESSION AW702855
VERSION AW702855.1 GI:7587023
KEYWORDS EST.
SOURCE Toxoplasma gondii.
ORGANISM Toxoplasma gondii.
REFERENCE Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.
AUTHORS Hehl,A., Manger,I., Marra,M., Sibley,L.D., Ajioke,J.A.,
Aslett,M.A., Dietrich,N., Dubuque,T., Hillier,L., Kucaba,T.,
Wan,K.L., Waterston,R.H. and Boothroyd,J.
TITLE WashU-Merck-Stanford-NIH Toxoplasma EST project
JOURNAL Unpublished (1996)
COMMENT On Jul 7, 1999 this sequence version replaced gi:5405402.
Contact: Maria M
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@watson.wustl.edu
Contact David Sibley (toxost@borcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gbco
High quality sequence stop: 416.

FEATURES
source

1. .451
Location/Qualifiers
/organism="Toxoplasma gondii"
/strain="VEG (Type III)"
/db_xref="taxon:5811"
/clone="TgESTz289f03.y1"
/clone_lib="TgVEG-Tachyzoite cDNA"
/dev_stage="Tachyzoite"
/lab_host="SOLR Cells"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was synthesized from poly mRNA using an
oligo-dT primer containing an XhoI site. Following second
strand synthesis, EcoRI adapters were ligated to the cDNA
and products were size- selected on sephacryl S500. The
cDNAs were ligated to EcoRI/ XhoI prepared lambda ZapII
(Stratagene). Clones were converted to phagemids by mass
excision using ExAssist helper phage and E.coli SOLR cells
(Stratagene). Insert sizes range from 0.3-3.0 kb. The
library may contain a small percentage of host or
bacterial contaminants. NOTE: Many clones lack the 5'
EcoRI site but contain the upstream polylinker sites."

BASE COUNT 130 a 127 c 137 g 57 t
ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x AW702855 ..
Align seg 1/1 to: AW702855 from: 1 to: 451

17 ArgProLysThrGluArgArgThr 24
|||||
101 AGCCGAAACGGAAGCGAACG 124

seq_name: gb_gss3:AQ237040

seq_documentation_block:
LOCUS AQ237040 583 bp DNA GSS 21-APR-1999
DEFINITION RPC111-71J1.TK RPCI-11 Homo sapiens genomic clone RPCI-11-
71J1, genomic survey sequence.
ACCESSION AQ237040

VERSION AQ237040.1 GI:3669331
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT On Sep 10, 1998 this sequence version replaced gi:3553918.
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.

FEATURES
source

1..583
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:7527096"
/db_xref="taxon:9606"
/clone="RPCI-11-71J1"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"

BASE COUNT 182 a 150 c 101 g 149 t
ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x AQ237040 ..
Align seg 1/1 to: AQ237040 from: 1 to: 583

46 ValLeuLeuCysGlyAspGluAla 53
|||||
403 GTTTTACTGTCGGTGACGAGCA 426

```

seq_name: gb_gss11:AQ753213
seq_documentation_block:
LOCUS      583 bp      DNA      22-JUL-1999
DEFINITION HS_5308_A2_H06_47A_RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate-884 Col-12 Row-O, genomic survey sequence.
ACCESSION  AQ753213

VERSION    AQ753213.1  GI:5578264
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 583)
            Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL    Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE    99380589
COMMENT    On Dec 15, 1999 this sequence version replaced gi:4212675.
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@redjong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 884 row: 0 column: 12
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 583.
FEATURES   Location/Qualifiers
            source
            1..583
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="RPCI-11 Human Male BAC Library"
            /sex="male"
            /note="vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
            Male blood DNA was isolated from one randomly chosen donor
            and partially digested with a combination of EcoRI and
            EcoRI Methylase. Size selected DNA was cloned into the
            pBACE3.6 vector at EcoRI sites"
BASE COUNT 151 a 152 c 132 g 138 t 10 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x AQ753213
..
Align seg 1/1 to: AQ753213 from: 1 to: 583
57 LysSerAlaThrIleuArgAlga 64
|||||
536 AAAAGTGCTACCTTGAGCGTGT 559

seq_name: gb_gss15:A2073363
seq_documentation_block:
LOCUS      672 bp      DNA      31-MAR-2000
DEFINITION RPCI-23-393022.TJB RPCI-23 Mus musculus genomic clone RPCI-23-
            393022, genomic survey sequence.
ACCESSION  A2073363

VERSION    A2073363.1  GI:7366260
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 672)
            Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S.,
            Akinret,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de
            Jong,P. and and Fraser,C.M.
            Mouse BAC End sequences from Library RPCI-23
            Unpublished (1999)
            On Mar 23, 1999 this sequence version replaced gi:3323682.
            Other_GSSs: RPCI-23-393022.TJ RPCI-23-393022.TVB
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@redjong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 393 row: 0 column: 22
            Seq primer: SP6
            Class: BAC ends.
FEATURES   Location/Qualifiers
            source
            1..672
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="RPCI-23-393022"
            /clone_lib="RPCI-23"
            /sex="Female"
            /lab_host="DH10B"
            /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
            EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
            brain genomic DNA was isolated and partially digested
            with a combination of EcoRI and EcoRI Methylase. Size
            selected DNA was cloned into the pBACE3.6 vector at the
            EcoRI sites. The ligation products were transformed into
            DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 263 a 135 c 109 g 164 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x A2073363
..
Align seg 1/1 to: A2073363 from: 1 to: 672
37 IleAsnAspArgIleGlnGlnLeu 44
|||||
42 ATAAATGACGAGATTCAACAGCTG 65

seq_name: gb_gss15:A2048465
seq_documentation_block:

```

LOCUS A2048465 1401 bp DNA GSS 17-MAR-2000
DEFINITION PSB141 Barley PstI genomic clones Hordeum vulgare genomic
clone PSB141, genomic survey sequence.
ACCESSION A2048465

VERSION A2048465.1 GI:7261122
KEYWORDS GSS.
SOURCE Hordeum vulgare
ORGANISM Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.
REFERENCE 1 (bases 1 to 1401)
AUTHORS Reeves, R.G., Pratchett, N. and Laurie, D.A.
TITLE Genomic sequences from a barley PstI library
JOURNAL Unpublished (2000)
COMMENT On Dec 15, 1999 this sequence version replaced gi:4574979.
Contact: Laurie DA
Cereals Research Department
John Innes Centre
Norwich Research Park, Colney, Norwich NR4 7UH, UK
Email: david.laurie@bbsrc.ac.uk
Mapped on chromosome 2HL.
PCR Primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Insert Length: 1401 Std Error: 0.00
Class: Shotgun
High quality sequence stop: 1401.
Location/Qualifiers
1..1401
/organism="Hordeum vulgare"
/cultivar="Betzes"
/db_xref="taxon:4513"
/clone="PSB141"
/clone_lib="Barley PstI genomic clones"
/sex="hermaphrodite"
/tissue_type="Leaf"
/note="Vector: PUC18; Site.1: PstI; Site.2: PstI; Genomic
fragments from PstI digest of DNA extracted from leaves."

BASE COUNT 385 a 342 c 358 g 316 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x A2048465/rev ..
Align seg 1/1 to reverse of: A2048465 from: 1 to: 1401

56 SerLysSerAlaThrLeuArgArg 63
|||||
138 AGCAATCAGCTACTTTGAGAAGA 115

seq_name: gb_est13:AA891352

seq_documentation_block:
LOCUS AA891352 126 bp mRNA EST 16-JUN-1998
DEFINITION EST195155 Normalized rat heart, Bento Soares Rattus sp. cDNA clone
RHAS96 3' end, mRNA sequence.
ACCESSION AA891352
VERSION AA891352.1 GI:3018231
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 126)
AUTHORS Lee,N.H., Glodok.A., Chandra,I., Mason,T.M., Quackenbush,J.,

Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
Gene Index
Unpublished (1998)
On Jul 19, 1995 this sequence version replaced gi:2938884.
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1..126
/organism="Rattus sp."
/db_xref="ATCC (inhost):2012057"
/db_xref="taxon:10118"
/clone="RHEAS96"
/clone_lib="Normalized rat heart, Bento Soares"
/note="Organ: heart; Vector: pT73Pac; Site.1: EcoRI;
Site.2: NotI"
26 a 15 c 18 g 67 t
BASE COUNT
ORIGIN

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x AA891352/rev ..
Align seg 1/1 to reverse of: AA891352 from: 1 to: 126

27 AsnLeuIleGluLysLysTyr 33
|||||
72 ATCTGATAGAGAAAGAT 52

seq_name: gb_est8:AA474206

seq_documentation_block:
LOCUS AA474206 148 bp mRNA EST 18-JUN-1997
DEFINITION ve54g10.r1 Beddington mouse embryonic region Mus musculus cDNA
clone IMAGE:822018 5' similar to SW:RL2B_HUMAN P39024 60S RIBOSOMAL
PROTEIN L23A ;, mRNA sequence.
ACCESSION AA474206
VERSION AA474206.1 GI:2202433
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 148)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HM Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HM Mouse EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:490298
Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand
Seq primer: -28m13 rev1 ET from AmerSham
High quality sequence stop: 1.

FEATURES

Location/Qualifiers
1. 148
/organism="Mus musculus"
/strain="C57BL6 x DBA"
/db_xref="taxon:10090"
/clone="IMAGE:822018"
/clone_lib="Beddington mouse embryonic region"
/sex="pooled"
/tissue_type="embryo"
/dev_stage="7.5dpc"
/lab_host="DH12S"

/note="organ: whole embryo; Vector: PCMV-SPORT; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Gastrulating embryos were collected at 7.5dpc from C57BL6 x DBA matings, excluding embryos that had developed head folds and all extraembryonic tissues. Average insert size: 1.3 kb (range: 0.5 - 3.0 kb). Referenced in Development 121, 2479-2489 (1995)."

BASE COUNT 61 a 36 c 40 g 10 t 1 others

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x AA474206 ..

Align seg 1/1 to: AA474206 from: 1 to: 148

16 ArgArgProLysThrGluArg 22
|||||
128 CGACGGCCCAAGCGACGCT 148

seq_name: gb_est31:AV164024

seq_documentation_block:

LOCUS AV164024 197 bp mRNA EST 08-JUL-1999
DEFINITION AV164024 Mus musculus head C57BL/6J 13-day embryo Mus musculus cDNA clone 3110018104, mRNA sequence.

ACCESSION AV164024
VERSION AV164024.1 GI:5430000

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 197)

REFERENCE

Akaiwa, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T.,

Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M.,

Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsuma, H., Oda, H.,

Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y.,

Sugahara, Y., Suzuki, H., Suzuki, H., Tatenno, M., Tomaru, Y.,

Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T.,

Yoshino, M., Muranatsu, M., Okazaki, Y. and Hayashizaki, Y.

Riken Mouse ESTs

Unpublished (1999)

On Feb 18, 1999 this sequence version replaced gi:4299503.

Contact: Chie Owa

Genome Science Laboratory

Riken

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp

Thermostabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length cDNA

(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES

source

Location/Qualifiers
1. 197
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="3110018104"
/clone_lib="Mus musculus head C57BL/6J 13-day embryo"
/sex="mixed"
/tissue_type="head"
/dev_stage="13-day embryo"
64 a 39 c 43 g 51 t

BASE COUNT

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x AV164024/rev ..

Align seg 1/1 to reverse of: AV164024 from: 1 to: 197

43 GlnLeuLysValLeuLeuCys 49

|||||

26 CAGCTGAAGTGCTTCTGTGT 6

seq_name: gb_est28:AJ282728

seq_documentation_block:

LOCUS AJ282728 199 bp mRNA EST 07-FEB-2000
DEFINITION 4A3A-ABB-D-05-R Anopheles gambiae immune competent 4A3A Anopheles gambiae cDNA clone 4A3A-ABB-D-05, mRNA sequence.

ACCESSION AJ282728

VERSION AJ282728.1 GI:6930607

KEYWORDS

EST.

SOURCE

African malaria mosquito.

ORGANISM

Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;

Culicoidae; Culicidae; Anopheles.

1 (bases 1 to 199)

Dimopoulos, G., Casavant, T.L., Chang, S., Scheetz, T., Roberts, C.,

Donohue, M., Schultz, J., Benes, V., Bork, P., Ansorge, W., Soares, M.B.

and Kafatos, F.C.

Anopheles gambiae pilot gene discovery project: Identification of

novel mosquito innate immunity genes from ESTs generated from

immune competent cell lines

Unpublished (2000)

Contact: Dimopoulos G

Fotis C. Kafatos Laboratory

European Molecular Biology Laboratory

Meyerhofstrasse 1, 69117 Heidelberg, Germany.

Location/Qualifiers

1. 199

/organism="Anopheles gambiae"

/strain="4A r/r"

/db_xref="taxon:7165"

/clone="4A3A-ABB-D-05"

/clone_lib="Anopheles gambiae immune competent 4A3A"

/cell_line="immune competent 4A3A"

/lab_host="E. coli DH10b"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: EcoRI; Site_2: NotI; sequenced from

forward priming site which reads from the 3' end of the

cDNA. The 4A3A is a directionally cloned and normalized

cDNA library that was constructed from the 4A3A cell line

oligo-T primed cDNA according to: Bonaldo, Lennon & Soares (1996) : Normalization and Subtraction: Two approaches To Facilitate Gene Discovery, Genome Research 6, 791-806."

BASE COUNT
ORIGIN

42 a 64 c 44 g 49 t

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x AJ282728/rev ..
Align seg 1/1 to reverse of: AJ282728 from: 1 to: 199

20 ThrGluArgArgThrAlaHis 26
|||||
78 ACTGAACGGCGCACCGCGCAC 58

seq_name: gb_est16:AI100958

seq_documentation_block:
LOCUS AI100958 206 bp mRNA EST 31-JAN-1999
DEFINITION EST210247 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
RBRB46 3' end, mRNA sequence.

ACCESSION AI100958
VERSION AI100958.1 GI:3705969
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 206)
AUTHORS Lee N.H., Glodek A., Chandra I., Mason T.M., Quackenbush J., Kriavage A.R. and Adams M.D.

TITLE Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat Gene Index

JOURNAL Unpublished (1998)
COMMENT Other ESTs: TC52573
Contact: Lee, NH
ATCC

The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES
source
1..206
Location/Qualifiers
/organism="Rattus sp."
/db_xref="ATCC (inhost):2029211"
/db_xref="taxon:10118"
/clone="RBRB46"
/clone_lib="Normalized rat brain, Bento Soares"
/note="Organ: brain; Vector: pT7T3Pac; Site_1: ECORI;
Site_2: NotI"

BASE COUNT 60 a 40 c 43 g 63 t
ORIGIN

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x AI100958/rev ..
Align seg 1/1 to reverse of: AI100958 from: 1 to: 206

27 AsnLeuIleGluLysLysTyr 33
|||||
47 AATCTGATAGAGAAAAAGTAT 27
seq_name: gb_est30:AV067361

seq_documentation_block:
LOCUS AV067361 206 bp mRNA EST 24-JUN-1999
DEFINITION AV067361 Mus musculus small intestine C57BL/6J adult Mus musculus
cDNA clone 2010204N08, mRNA sequence.

ACCESSION AV067361
VERSION AV067361.1 GI:5187189
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 206)

REFERENCE 1 (bases 1 to 206)
AUTHORS Carninci P., Shibata K., Ozawa Y., Konno H., Itoh M., Aizawa K., Akahira S., Akiyama J., Fukuda S., Fukunishi Y., Funayama T., Hara A., Hayatsu N., Hori F., Ishikawa T., Itoh M., Izawa M., Kawai J., Kikuchi N., Kojima Y., Matsuyama T., Niitsuma H., Oda H., Owa C., Sato K., Shibata Y., Shigemoto Y., Shiraki T., Sogabe Y., Sugahara Y., Suzuki H., Tateo M., Tomaru Y., Tomimaga N., Watanabe S., Yagane M., Yamamura T., Yokota T., Yoshino M., Muramatsu M., Okazaki Y. and Hayashizaki Y.

TITLE Unpublished (1999)
JOURNAL RIKEN Mouse ESTs
COMMENT On Jun 15, 1998 this sequence version replaced gi:3224182.
Contact: Chie Owa
Genome Science Laboratory
RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES
Location/Qualifiers
1..206
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2010204N08"
/clone_lib="Mus musculus small intestine C57BL/6J adult"
/sex="male"
/tissue_type="small intestine"
/dev_stage="adult"

BASE COUNT 86 a 23 c 26 g 71 t
ORIGIN

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x AV067361/rev ..
Align seg 1/1 to reverse of: AV067361 from: 1 to: 206

43 GlnLeuLysValLeuLeuCys 49
|||||
118 CAATTGAAAGTATTATTATGT 98

seq_name: gb_est33:AI676373

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus; 1 (Bases 1 to 224)

Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Ishii,Y., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Kawai,C., Kawai,J., Kikuchi,N., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigenoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateo,M., Tomihaga,N., Tsunoda,Y., Watabiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y. Riken Mouse ES Cells (Konno,H., et al.)
 Unpublished (1999)

On Jul 9, 1999 this sequence version replaced gi:5435187.
 Contact: Yoshihide Hayashizaki
 Genome Exploration Research Group, Life Science Tsukuba Center,
 The Institute of Physical and Chemical Research (RIKEN), Genomic
 Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: +81-298-36-9013
 Fax: +81-298-36-9098
 Email: genome-res@rtc.riken.go.jp,
 URL:http://genome.rtc.riken.go.jp/
 Sasaki,N., Izawa,M., Watabiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
 Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
 Itoh,M., Kitsuunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

Location/Qualifiers
 1..224
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="5530400N10"
 /clone_lib="RIKEN full-length enriched, 10 days neonate head"
 /sex="mixed"
 /tissue_type="head"
 /dev_stage="10 days neonate"
 /lab_host="DH10B"

/note="Site_1: Sali; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGGATCCAGAGCTCTTTTATTTTATTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGATTAAATTAATCCGCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end: BamHI."

73 a 45 c 48 g 58 t

BASE COUNT
 ORIGIN

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x AV293889 ..

Align seg 1/1 to: AV293889 from: 1 to: 224

16 ArgArgProLysThrIluArg 22

|||||

155 AGGAGACCTAAACCGAGAGG 175

seq_name: gb_est3:AA199336

seq_documentation_block:
 LOCUS AA199336 237 bp mRNA EST 15-FEB-1997
 DEFINITION mt55f04.r1 Stragatene mouse embryonic carcinoma (#937317) Mus musculus cDNA clone IMAGE:633823 5', mRNA sequence.

ACCESSION AA199336

VERSION AA199336.1 GI:1794957

KEYWORDS

SOURCE EST.

ORGANISM house mouse.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 237)

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:385815

Seq primer: -28ml3 rev1 ET from Amersham

High quality sequence stop: 134.

FEATURES

source

1..237

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:633823"

/clone_lib="Stragatene mouse embryonic carcinoma

(#937317)"

/tissue_type="carcinoma"

/dev_stage="embryonic"

/lab_host="SOLR (kanamycin resistant)"

/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:

XhoI; Cloned unidirectionally. Primer: Oligo dT. P19 cell

line. Cloned insert size: 1.0 kb; Uni-ZAP XR Vector: -5'

adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor

sequence: 5' CTGAGTGTGTGTGTGTGTGT 3'

57 a 52 c 46 g 82 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 7.00 Length: 7

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x AA199336 ..

Align seg 1/1 to: AA199336 from: 1 to: 237

44 IeulysValLeuLeuCysGly 50

|||||

6 TTTAAAGCTTTTACTCTGTGGC 26

seq_name: gb_est45:AW641622

seq_documentation_block:

LOCUS AW641622 240 bp mRNA EST 03-APR-2000

DEFINITION cm09d12.w1 Blackshear/Soares normalized Xenopus egg library Xenopus laevis cDNA clone PBX0108D12 5', mRNA sequence.

ACCESSION AW641622

VERSION AW641622.1 GI:7398890

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM

Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 240)

Blackshear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G.

Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M.,

Touchman, J.W., Bonaldo, M.F., and Soares, M.B.

The NIEHS Xenopus Maternal EST Project

Unpublished (2000)

On Jan 6, 2000 this sequence version replaced gi:6677531.

Contact: Perry J. Blackshear

Office of Clinical Research and Laboratory of Signal Transduction

National Institute of Environmental Health Sciences

A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,

USA

Tel: 919 541-4899

Fax: 919 541-4571

Email: black009@niehs.nih.gov

Clone is available through Research Genetics, Inc., 2130 Memorial

Parkway, Huntsville, AL 35901

phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email

cdna@resgen.com

DNA Sequencing and analyses performed by National Institutes of

Health Intramural Sequencing Center (NISC).

PCR Primers

FORWARD: TGTAAACGACGCCGAGT

BACKWARD: CAGGAACAGCTATGACC

Plate: 0108 row: D column: 12

Seq primer: T7 primer.

FEATURES

source

1..240

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone="PBX0108D12"

/clone_lib="Blackshear/Soares normalized Xenopus egg

library"

/sex="female"

/tissue_type="unfertilized egg"

/cell_type="unfertilized egg"

/dev_stage="unfertilized egg"

/lab_host="PH10B"

/note="Vector: pT7T3-Pac; Site.1: EcoRI; Site.2: NotI;

PolyA-selected mRNA was prepared from unfertilized Xenopus

laevis eggs. The library was constructed in the vector

pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and

Soares, M.B. 'Normalization and subtraction: two

approaches to facilitate gene discovery', Genome Research

6:791-806, 1996. The first strand synthesis used a

NotI-dT18 primer; double stranded cDNAs were ligated to

EcoRI adaptors, digested with NotI, and directionally

cloned into the NotI and EcoRI-digested pT7T3-Pac vector.

The library contained approximately 7.2 X 10⁵

recombinants, with average insert sizes of 1-1.5 kb."

69 a 40 c 60 g 71 t

BASE COUNT

ORIGIN

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x AW641622/rev ..

Align seg 1/1 to reverse of: AW641622 from: 1 to: 240

41 IleGlnGlnLeuLysValLeu 47
 |||||
 166 ATACAGCAGCTTAAGTTCTT 146

seq_name: gb_est26:AI897499

seq_documentation_block:

LOCUS AI897499 246 bp mRNA EST 27-JUL-1999
 DEFINITION EST266942 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
 CLED27D12, mRNA sequence.
 ACCESSION AI897499
 VERSION AI897499.1 GI:5603401
 KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
 Solanaceae; Solanum; Lycopersicon.
 REFERENCE 1 (bases 1 to 246)
 AUTHORS Alcalá, J., Vrebalov, J., White, R., Matern, A.L., Vision, T.,
 Holt, I.E., Liang, F., Upton, J., Ronning, C.M., Craven, M.B.,
 Fujii, C.Y., Bowman, C.L., Nierman, W., Fraser, C.M., Venter, J.C.,
 Martin, G.B., Tanksley, S.D. and Giovannoni, J.
 TITLE Generation of ESTs from tomato carpel tissue
 JOURNAL Unpublished (1999)
 COMMENT On Nov 2, 1998 this sequence version replaced gi:3829079.
 Contact: David Frisch
 Clemson University
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 4366
 Fax: 864 656 4293
 Email: dirisch@clemson.edu
 5 prime sequence.

FEATURES

Source
 1..246
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="CLED27D12"
 /clone_lib="tomato ovary, TAMU"
 /tissue_type="carpel"
 /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
 /lab_host="Xli-Blue MRF"
 /note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
 directionally cloned cDNA in vector Lambda ZAP II with 5'
 and 3' ends located at the EcoRI and XhoI sites,
 respectively."

BASE COUNT 66 a 42 c 62 g 76 t
 ORIGIN

alignment_scores:

Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x AI897499 ..

Align seg 1/1 to: AI897499 from: 1 to: 246

44 LeuLysValLeuLeuCysGly 50
 |||||
 108 TTGAGGTGCTGCTATGTGGT 128

seq_name: gb_gss18:FR0013555

seq_documentation_block:

LOCUS FR0013555 248 bp DNA GSS 18-SEP-1997
 DEFINITION F.rubripes GSS sequence, clone 125H09ac8, genomic survey sequence.
 ACCESSION AL004804
 VERSION AL004804.1 GI:2450374
 KEYWORDS GSS; genome survey sequence.
 SOURCE Fugu rubripes.
 ORGANISM Fugu rubripes.
 Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
 Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
 Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.

REFERENCE

1 (bases 1 to 248)

AUTHORS

Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrana, Y.,
 Williams, G. and Brenner, S.

TITLE

Direct Submission

JOURNAL

Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource
 Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgm.mrc.ac.uk

COMMENT

Vector: pBluescript II KS
 V type: phagemid
 PRIMER: KS
 DESC: One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

FEATURES

Location/Qualifiers

1..248

/organism="Fugu rubripes"

/db_xref="taxon:31033"

/clone_lib="cosmid 125H09"

/clone="125H09ac8"

BASE COUNT 58 a 51 c 64 g 70 t 5 others

ORIGIN

alignment_scores:

Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x FR0013555 ..

Align seg 1/1 to: FR0013555 from: 1 to: 248

45 LysValLeuLeuCysGlyAsp 51
 |||||
 177 AAAGTCTGCTGTGTGGGAC 197

seq_name: gb_est33:AV274480

seq_documentation_block:

LOCUS AV274480 249 bp mRNA EST 05-NOV-1999
 DEFINITION AV274480 RKEN full-length enriched, adult male testis (DH10B) Mus
 musculus cDNA clone 4932418P15 3', mRNA sequence.

ACCESSION

AV274480

VERSION

AV274480.1 GI:6262517

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 249)

AUTHORS

Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,
 Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,
 Ishii, Y., Ishikawa, T., Iton, M., Izawa, M., Kadota, K., Kagawa, I.,

Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K.,
Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Suganara, Y.,
Suzuki, H., Suzuki, H., Takahashi, F., Tateono, M., Tomimaga, N.,
Tsunoda, Y., Wachihi, A., Wakatsuki, S., Yamamura, T., Yasunishi, A.,
Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Riken Mouse ESTs (Konno, H., et al.)
Unpublished (1999).

FEATURES

```

Location/Qualifiers
1. 249
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4932418P15"
/clone_lib="RIKEN full-length enriched, adult male testis
(DH10B)"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; CDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGAGATCCAGAGCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalase thermo-activated reverse
transcriptase and subsequently enriched for full-length b
cap-trapper. Second strand cDNA was prepared with the
primer adaptor of sequence [5',
GAGAGAGAGATCTCGAGTTAATTAAATATCCCCCCCCCCCCC 3']. cDNA
was cloned into the XhoI and BamHI sites. Vector: a
modified pBluescript KS(+) after bulk excision from Lambd
FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
a 47 c 46 g 76 t

```

```

BASE COUNT      80 a    47 c    46 g    76 t
ORIGIN

alignment_scores:
    Quality:      7.00      Length:      7
    Ratio:        1.000     Gaps:        0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:

```

```

US-09-332-522B-2_COPY_335_428 x AV274480 ..
Align seg 1/1 to: AV274480 from: 1 to: 249

52 GluAlaLysLeuSerIysSer 58
| | | | | | | | | | | | | |
33 GAAGCAAAGCTCTCAAAAAGT 53

seq_name: gb_est53:W14668

seq_documentation_block:
LOCUS W14668 259 bp mRNA EST 10-SEP-1996
DEFINITION mb30f02.r1 Soares mouse p3BNMF19.5 Mus musculus cDNA clone
IMAGE:330939 5', mRNA sequence.

```

FEATURES

```

1. 259
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:330939"
/clone.lib="Soares mouse p3NNF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTCGGAGCGCGCATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fátima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
56 c 60 g 83 t
60 a

```

```

BASE COUNT      60 a   56 c   60 g   83 t
ORIGIN
          DR. MINORU AO (Wayne State University)

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x W14668 ..

```


alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x AW332840/rev ..

Align seg 1/1 to reverse of: AW332840 from: 1 to: 265

27 AsnLeuileGluLysLysTyr 33
 |||||
 40 AATCTGATAGAGAAAAAGTAT 20

seq_name: gb_gss14:AZ002831

seq_documentation_block:
 LOCUS AZ002831 270 bp DNA GSS 24-FEB-2000
 DEFINITION RPI-23-344L19.TJ RPI-23 Mus musculus genomic clone RPI-23-344L19, genomic survey sequence.
 ACCESSION AZ002831

VERSION AZ002831.1 GI:7078187
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 270)
 Zhaio S., Nierman W., Feldblyum T., Malek J., Shatsman S., Akintet B., Levins M., McGann S., Tsengaye G., Geer K., Krol M., de Jong P. and Fraser C.M.
 Mouse BAC End Sequences from Library RPI-23
 Unpublished (1999)
 On Feb 19, 1999 this sequence version replaced gi:4145974.
 Other_GSSs: RPI-23-344L19.TV

CONTACT: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: shao@tigr.org

Clones are derived from the mouse BAC library RPI-23. For BAC library availability, please contact Pieter de Jong (pieter@tigr.org, med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 344 row: L column: 19
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source
 Location/Qualifiers
 1..270
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPI-23-344L19"
 /clone_lib="RPI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1: EcoRI; Site: 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI. Methyase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT
 105 a 49 c 57 g 59 t

ORIGIN

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x AZ002831 ..

Align seg 1/1 to: AZ002831 from: 1 to: 270

20 ThrGluArgArgThrAlaHis 26
 |||||
 124 ACAGAACGACGACTGCACAT 144

seq_name: gb_est46:C17181

seq_documentation_block:
 LOCUS C17181 272 bp mRNA EST 30-SEP-1996
 DEFINITION C17181 Clontech human aorta polyA+ mRNA (#6572) Homo sapiens CDNA clone GEN-541B07 5', mRNA sequence.
 ACCESSION C17181
 VERSION C17181.1 GI:1571888
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 272)
 FUJIWARA T., Hirano H., Katagiri T., Kawai A., Kuga Y., Nagata M., Okuno S., Ozaki K., Shimizu F., Shimada Y., Shinomiya H., Takachi A., Takeda S., Watanabe T., Takahashi E., Hirai Y., Maekawa H., Shin S. and Nakamura Y.
 Fujiwara et al. (1995)
 Unpublished (1995)
 On Jul 7, 1999 this sequence version replaced gi:5408920.
 CONTACT: Tsutomu Fujiwara
 Otsuka GEN Research Institute
 Otsuka Pharmaceutical Co., Ltd
 463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
 Tel: 0886-65-2888
 Fax: 0886-37-1035.

FEATURES
 source
 Location/Qualifiers
 1..272
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="GEN-541B07"
 /clone_lib="Clontech human aorta polyA+ mRNA (#6572)"

BASE COUNT
 73 a 68 c 86 g 45 t

ORIGIN

46 ValLeuLeuCysGlyAspGlu 52
 |||||
 73 GTTCTCTCTGTGGAGATGAG 53

seq_name: gb_est49:H65208

seq_documentation_block:
 LOCUS H65208 272 bp mRNA EST 11-DEC-1995
 DEFINITION YU66f02.s1 Weizmann Olfactory Epithelium Homo sapiens CDNA clone IMAGE:238779 3' similar to contains Alu repetitive element; mRNA sequence.

Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y.,
 Sebastiani-Kabaktchis, C. and Tessier, A. Analysis of the human genome
 IMAGE: molecular integration of the analysis of the human genome
 and its expression
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
 95277534
 On Sep 12, 1996 this sequence version replaced gi:1397804.
 Contact: Genethon
 Genexpress-Genethon
 Genethon Centre de recherche sur le Genome Humain
 1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698
 Email: genexpress@genethon.fr
 Single read.
 Genethon
 Genexpress_library_id: C; Genexpress_sequence_id: y1c-0gg12
 Seq primer: (-21)M13.universal.

FEATURES

source
 1. .291
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="c-0gg12"
 /clone_lib="normalized infant brain cDNA"
 /sex="Female"
 /tissue_type="total brain"
 /dev_stage="3 months old"
 /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
 Site_2: NotI; sex: Female; dev_stage=3 months old;
 isolate=muscular atrophy patient; tissue_type=total
 brain; total mRNA was oligo-(dT) primed and directionally
 cloned 5' -> 3' into the HindIII -> NotI sites of the
 lafmid BA vector. Clone library from B. Soares, Psychiatry
 Dept. Columbia University, USA. Normalization method:
 Banto Soares, P.N.A.S in press"
 BASE COUNT 67 a 68 c 81 t 1 others
 ORIGIN

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x Z42367 ..
 Align seg 1/1 to: Z42367 from: 1 to: 291

44 LeuLysValLeuLeuCysGly 50
 |||||
 22 TTAAGGTCTCTTTCGGGT 42

seq_name: gb_est16:AI106092

seq_documentation_block:
 LOCUS AI106092 296 bp mRNA EST 25-AUG-1998
 DEFINITION cb02f06_t3 2F adult heart library Danio rerio cDNA 5 prime, mRNA
 sequence.

ACCESSION AI106092.1 GI:3461195
 VERSION AI106092
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Cyprinidae; Rasbora; Danio.
 1 (bases 1 to 296)
 Chen, J.N., Desauvage, F., Hosobuchi, M., Jackson, D.G. and
 Fishman, M.C.

Expressed Sequences from The Adult Zebrafish Heart
 JOURNAL Unpublished (1998)
 COMMENT On Apr 7, 1998 this sequence version replaced gi:3035866.
 Contact: Mark C. Fishman

Cardiovascular Research Center
 Massachusetts General Hospital
 Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA
 Fax: 6177265806
 Email: fishman@mgm.cvr.harvard.edu
 http://zebrafish.mgh.harvard.edu
 The original clones used for sequencing are no longer available;
 the library is available from Mark C. Fishman.
 Insert length: 296 Std Error: 0.00
 Seq primer: t3.

FEATURES

source
 Location/Qualifiers
 1. .296
 /organism="Danio rerio"
 /strain="AB"
 /db_xref="taxon:7955"
 /clone_lib="ZF adult heart library"
 /sex="mixed"
 /tissue_type="myocardium, endocardium, vessel"
 /dev_stage="adult"
 /lab_host="E. coli XL1 Blue"
 /note="Organ: heart; Vector: LambdaZAPII; Site_1: EcoRI;
 Site_2: XhoI"
 BASE COUNT 85 a 72 c 88 g 45 t 6 others
 ORIGIN

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2_COPY_335_428 x AI106092 ..
 Align seg 1/1 to: AI106092 from: 1 to: 296

57 LysSerAlaThrLeuArgArg 63
 |||||
 6 AAGAGCGCAACCTCCGCGC 26

seq_name: gb_est12:AA786770

seq_documentation_block:
 LOCUS AA786770 300 bp mRNA EST 31-JUL-1998
 DEFINITION msh12a1.r1 Aspergillus nidulans 24hr asexual developmental and
 vegetative cDNA lambda zap library Aspergillus nidulans cDNA clone
 msh12a1 5', mRNA sequence.

ACCESSION AA786770
 VERSION AA786770.1 GI:2847100
 KEYWORDS EST.
 SOURCE Aspergillus nidulans.
 ORGANISM Aspergillus nidulans

Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae;
 anamorphic Trichocomaceae; Aspergillus.
 1 (bases 1 to 300)
 Kuper, D., Gray, J., Hausner, J., Lai, H., Martin, W., Aramayo, R.,
 Prade, R. and Roe, B.

An Aspergillus nidulans EST Database
 JOURNAL Unpublished (1998)
 COMMENT On Oct 31, 1997 this sequence version replaced gi:1520944.
 Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal
 Genetics Stock Center
 Seq primer: SK
 High quality sequence stop: 234.
 Location/Qualifiers
 1. .300

FEATURES

source
 Location/Qualifiers
 1. .300

```

/organism="Aspergillus nidulans"
/strain="FGSC A28"
/db_xref="taxon:5072"
/clone="m3h12a1"
/clone_lib="Aspergillus nidulans 24hr asexual
developmental and vegetative cDNA lambda zap library"
/tissue_type="vegetative mycelia, asexual structures"
/notes="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT      82 a      114 c      47 g      57 t
ORIGIN

alignment_scores:
  Quality:      7.00      Length:      7
  Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x AA786770 .
Align seg 1/1 to: AA786770 from: 1 to: 300

61 LeuAtgAqAlaIleGluHis 67
|||||
74 TTAAGACGGCCATCGAAT 94

seq_name: gb_est46:C30348

seq_documentation_block:
LOCUS      C30348      300 bp      mRNA      EST      18-OCT-1999
DEFINITION C30348 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
Caenorhabditis elegans cDNA clone yk246d4 3', mRNA sequence.
ACCESSION C30348
VERSION   C30348.1 GI:2362144
KEYWORDS EST.
SOURCE    Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
           Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
           Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 300)
AUTHORS   Kohara,Y., Mochizashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
           Sano,M., Miyata,A. and Nishigaki,A.
TITLE     Expression map of the C.elegans genome
JOURNAL   Unpublished (1996)
COMMENT   On Jan 6, 2000 this sequence version replaced gi:6673963.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES             Location/Qualifiers
     source            1..300
     organism="Caenorhabditis elegans"
     strain="N2"
     db_xref="taxon:6239"
     clone="yk246d4"
     clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"
     sex="hermaphrodite"
     tissue_type="embryo"
     dev_stage="embryo"

BASE COUNT      102 a      66 c      47 g      80 t      5 others
ORIGIN

alignment_scores:
  Quality:      7.00      Length:      7
  Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:
US-09-332-522B-2_COPY_335_428 x C30348/rev ..
Align seg 1/1 to reverse of: C30348 from: 1 to: 300

37 IleAsnAspArgIleGlnGln 43
|||||
82 ATAAATGACAGGATTCACAA 62

seq_name: gb_est47:C53448

seq_documentation_block:
LOCUS      C53448      300 bp      mRNA      EST      11-SEP-1997
DEFINITION C53448 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk318a7 3', mRNA sequence.
ACCESSION C53448
VERSION   C53448.1 GI:2391205
KEYWORDS EST.
SOURCE    Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
           Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
           Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 300)
AUTHORS   Kohara,Y., Mochizashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
           Sano,M., Miyata,A. and Nishigaki,A.
TITLE     Expression map of the C.elegans genome
JOURNAL   Unpublished (1996)
COMMENT   On Jan 6, 2000 this sequence version replaced gi:6677258.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES             Location/Qualifiers
     source            1..300
     organism="Caenorhabditis elegans"
     strain="C31489 him-8(e1489)"
     db_xref="taxon:6239"
     clone="yk318a7"
     clone_lib="Yuji Kohara unpublished cDNA"
     sex="hermaphrodite, male"
     tissue_type="whole animal"
     dev_stage="varied"

BASE COUNT      97 a      31 c      79 g      89 t      4 others
ORIGIN

alignment_scores:
  Quality:      7.00      Length:      7
  Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x C53448/rev ..
Align seg 1/1 to reverse of: C53448 from: 1 to: 300

26 HisAsnLeuIleGluIysLys 32
|||||
116 CATAAATTTAATAGAAAAAAA 96

seq_name: gb_est34:AV298105

seq_documentation_block:
LOCUS      AV298105      302 bp      mRNA      EST      10-NOV-1999
DEFINITION AV298105 RIKEN full-length enriched, 8 days embryo Mus musculus
cDNA clone 573045411 3' similar to S82462 GATA-6-zinc finger
transcription factor, mRNA sequence.
ACCESSION AV298105
VERSION   AV298105.1 GI:6330346
KEYWORDS EST.

```


Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x AA480552 ..

Align seg 1/1 to: AA480552 from: 1 to: 305

1 SerAspSerglygluThrMet 7

|||||
54 TCTGACAGTGGAGAACTATG 74

seq_name: gb_est31:AV145949

seq_documentation_block:
LOCUS AV145949 310 bp mRNA EST 03-JUL-1999
DEFINITION AV145949 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA
clone 2810449P12, mRNA sequence.

ACCESSION AV145949

VERSION AV145949.1 GI:5350082

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 310)

AUTHORS

Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T.,
Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M.,
Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsunuma, H., Oda, H.,
Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y.,
Sugahara, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y.,
Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T.,
Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

TITLE

RIKEN Mouse ESTs

Unpublished (1999)

JOURNAL

On Jun 22, 1998 this sequence version replaced gi:3246767.

COMMENT

Contact: Chie Owa

Genome Science Laboratory

RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp
Thermolabile and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES

Location/Qualifiers

1..310

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="2810449P12"

/clone.lib="Mus musculus C57BL/6J 10-11 day embryo"

/sex="mixed"

/dev_stages="10-11 day embryo"

64 a 73 c 87 g 86 t

BASE COUNT
ORIGIN

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x AV145949 ..

Align seg 1/1 to: AV145949 from: 1 to: 310

59 AlaThrLeuArgArgAlaIle 65

|||||

19 GCAACTCTTCGAAGGCCATC 39

seq_name: gb_est6:AA356000

seq_documentation_block:

LOCUS AA356000 312 bp mRNA EST

DEFINITION EST64653 Jurkat T-cells VI Homo sapiens cDNA 5' end, mRNA sequence.

ACCESSION AA356000

VERSION AA356000.1 GI:2008470

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 312)

AUTHORS

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,
Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,
White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C.,
Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,
Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S.,
Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,
Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
Moreno-Palao, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
Dimke, D., Feng, D.F., Ferlie, A., Fischer, C., Hastings, G.A.,
He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,
Kozak, D.L., Kunsch, C., Hung, J., Li, H., Weissner, P.S., Olsen, H.,
Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M.,
Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,
Fraser, C.W. and Venter, J.C.

Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)

12140200

Other ESTs: THG168944

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers

1..312

/organism="Homo sapiens"

/db_xref="ATCC (inhost):158285"

/db_xref="taxon:9606"

/clone.lib="Jurkat T-cells VI"

/cell_type="T-lymphocyte"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"

80 a 70 c 76 g 84 t 2 others

BASE COUNT
ORIGIN

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x AA356000 ..

Align seg 1/1 to: AA356000 from: 1 to: 312

44 LeuLysValLeuLeuCysGly 50
|||||
200 TTAAGGTTCTTCTTGGCGT 220

seq_name: gb_est13:AA900350

seq_documentation_block: 314 bp mRNA EST 05-FEB-1999
LOCUS AA900350
DEFINITION UI-R-E0-cp-h-01-0-UI-s2 UI-R-E0 Rattus norvegicus cDNA clone
UI-R-E0-cp-h-01-0-UI 3' similar to gi|349088|gb|L12459|RATLYSOZYM
Rat lysozyme gene exons 1-4, complete cds, mRNA sequence.

ACCESSION AA900350

VERSION AA900350.1

KEYWORDS GI:4232841

SOURCE EST.

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 314)

REFERENCE Bonaudo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL 97044477

COMMENT On Apr 7, 1998 this sequence version replaced gi:3035704.

Program for Rat Gene Discovery and Mapping

Contact: Soares, MB

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: mscoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the

oligo-dt track served to identify it as a clone from the normalized

adult 8-Day-Embryo library. cDNA Library Preparation: M. Fatima

Bonaudo, Ph.D. Clone distribution: clones will be available through

Research Genetics

Seq primer: M13 Forward.

FEATURES

Location/Qualifiers

1..314

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-E0-cp-h-01-0-UI"

/clone_lib="UI-R-E0"

/dev_stage="embryonic"

/lab_host="DH10B (Life Technologies)"

/note="vector: p773b-Pac (Pharmacia) with a modified

polylinker; Site_1: NotI; Site_2: EcoRI; This library

consists of a mixture of individually tagged normalized

libraries constructed from 8, 12 and 18-day embryo. The

tag is a string of 3-5 nucleotides present between the

Not I site and the oligo-dt track which allows

identification of the library of origin of a clone within

the mixture."

BASE COUNT 84 a 60 c 70 g 100 t

ORIGIN

alignment_scores:

Quality: 7.00 Length: 7

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x AA900350/rev ..

Align seg 1/1 to reverse of: AA900350 from: 1 to: 314

27 AsnLeuIleGluLysLysTyr 33

|||||
64 AATCTGATAGAGAAAAAGTAT 44

seq_name: gb_est53:W06635

seq_documentation_block: 316 bp mRNA EST 12-AUG-1996
LOCUS W06635
DEFINITION T2197 MVAT4 bloodstream form of serodeme WRATat1.1 Trypanosoma
brucei rhodesiense cDNA 5', mRNA sequence.

ACCESSION W06635

VERSION W06635.1

KEYWORDS GI:1280713

SOURCE EST.

ORGANISM

Trypanosoma brucei rhodesiense.

Trypanosoma brucei rhodesiense

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

REFERENCE 1 (bases 1 to 316)

AUTHORS Djikeng,A., Donelson,J.E. and Majiwa,P.A.O.

Generation of expressed sequence tags as physical landmarks in the

genome of Trypanosoma brucei

Unpublished (1996)

COMMENT On May 10, 1995 this sequence version replaced gi:805630.

Contact: Majiwa PAO

Molecular Biology Unit

International Livestock Research Institute

P.O. Box 30709, Nairobi, Kenya

Tel: 254-2 630743

Fax: 254-2 631499

Email: p.majiwa@cgnet.com

Seq primer: T3 primer.

FEATURES

Location/Qualifiers

1..316

/organism="trypanosoma brucei rhodesiense"

/db_xref="taxon:31286"

/clone_lib="MVAT4 bloodstream form of serodeme WRATat1.1"

/note="Vector: Lambda ZAP II (Stratagene); Site_1: EcoRI;

Site_2: XhoI; The mRNA was purified from a cloned

population of bloodstream trypanosomes reexpressing the

MVAT4 metacyclic variant surface glycoprotein (VSG). A

unidirectional oligo dt-primer EcoRI/XhoI cDNA library was

constructed in lambda ZAP II (Stratagene)."

BASE COUNT 90 a 63 c 73 g 90 t

ORIGIN

alignment_scores:

Quality: 7.00 Length: 7

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x W06635 ..

Align seg 1/1 to: W06635 from: 1 to: 316

58 SerAlaThrLeuArgArgAla 64

|||||

99 TCTGCAACCTTCGCGCGCA 119

seq_name: gb_est49:H64717

seq_documentation_block:

LOCUS H64717

DEFINITION yu66f02.r1 Weizmann Olfactory Epithelium Homo sapiens cDNA clone

IMAGE:238779 5' similar to contains Alu repetitive element;; mRNA

sequence.

ACCESSION H64717

VERSION H64717.1

KEYWORDS GI:1023457

SOURCE EST.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

```

REFERENCE
AUTHORS      1 (bases 1 to 317)
              Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
              Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
              Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
              Marais,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
              Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Neg,J.,
              Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
              and Marra,M.

TITLE
JOURNAL      Generation and analysis of 280,000 human expressed sequence tags
MEDLINE      Genome Res. 6 (9), 807-828 (1996)
COMMENT      On Nov 22, 1999 this sequence version replaced gi:6461229.
              Contact: Wilson RK
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              High quality sequence stops: 237
              Source: IMAGE Consortium, LLNL
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              Seq primer: M13Rpl
              High quality sequence stop: 237.

FEATURES
            source
            1..317
              /organism="Homo sapiens"
              /db_xref="GDB:3864163"
              /db_xref="taxon:9606"
              /clone="IMAGE:238779"
              /clone_lib="Weizmann Olfactory Epithelium"
              /sex="Female"
              /tissue_type="olfactory epithelium"
              /dev_stage="35 year old"
              /lab_host="SOLR cells (kanamycin resistant)"
              /note="Organ: nose; Vector: pBluescript SK-; Site:1:
              EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:
              Oligo dr: Olfactory epithelium, normal. Average insert
              size: 0.8 Kb; Uni-ZAP XR Vector. Library constructed by N.
              Walker, D. Lancet, Weizmann Institute of Science. -5'
              adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
              sequence: 5' CTCGAGTGTGTTTTTTTTTTT 3'"
              90 a 69 c 75 g 81 t 2 others

BASE COUNT
ORIGIN
alignment_scores:
  Quality: 7.00 Length: 7
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
  US-09-332-522B-2_COPY_335_428 x H64717/rev ..
  Align seg 1/1 to reverse of: H64717 from: 1 to: 317

32 LysTyrArgCysSerIleAsn 38
|||||
96 AAATACAGATGTAGCATCAAT 76

seq_name: gb_est54:W85333

seq_documentation_block:
LOCUS      W85333          322 bp mRNA EST 12-SEP-1996
DEFINITION m146d10.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
clone IMAGE:408115 5', mRNA sequence.
ACCESSION W85333
VERSION   W85333.1 GI:1538740
KEYWORDS  EST.
SOURCE    house mouse.
ORGANISM  Mus musculus
           Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS      1 (bases 1 to 322)
              Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
              Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
              Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
              Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
              Waterston,R.
              The WashU-HMMI Mouse EST Project
              Unpublished (1996)
              On Sep 12, 1996 this sequence version replaced gi:1397874.
              Contact: Marra M/Mouse EST Project
              WashU-HMMI Mouse EST Project
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: mouseest@watson.wustl.edu
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              MGI:251883
              Seq primer: -28M13 rev2 from Amersham
              High quality sequence stop: 139.

FEATURES
            source
            1..322
              /organism="Mus musculus"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone="IMAGE:408115"
              /clone_lib="Soares mouse embryo NbME13.5 14.5"
              /sex="unknown"
              /tissue_type="embryo"
              /dev_stage="13.5-14.5dpc total fetus"
              /lab_host="DH10B"
              /note="vector: pT7T3D-Pac (Pharmacia) with a modified
              polylinker; Site:1: Not I; Site:2: Eco RI; 1st strand cDNA
              was primed with a Not I - oligo(dT) primer [5',
              TGTACCAACTGAGTGGAGGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT
              T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
              14.5dpc embryos [total RNA provided by Minoru KO, Wayne
              State Univ., from 2 ]; double-stranded cDNA was ligated to
              Eco RI adaptors (Pharmacia), digested with Not I and
              cloned into the Not I and Eco RI sites of the modified
              pT7T3 vector. Library went through one round of
              normalization, and was constructed by Bento Soares and
              M.Fatima Bonaldo."
              61 a 104 c 94 g 63 t

BASE COUNT
ORIGIN
alignment_scores:
  Quality: 7.00 Length: 7
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
  US-09-332-522B-2_COPY_335_428 x W85333/rev ..
  Align seg 1/1 to reverse of: W85333 from: 1 to: 322

41 IleGlnGlnLeuLysValLeu 47
|||||
224 ATCCAGCAACTCAAGTCTC 204

```

THIS PAGE BLANK (USPTO)

OM of: US-09-332-522B-2 to: EST:* out-format : pfs

Date: Sep 27, 2000 1:46 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+g2n.model -DEV=xlp  
-O=/cgn2.1/USPTO_spool/US09332523/runat.27092000.091221.6524/app-query.fasta.1.1355  
-DB=EST -OFTM=fastap -SUFFIX=01i.rst -GAPOP=4.500 -GAPEXT=0.050  
-MINMATCH=0.100 -ICOPCL=0.000 -ICOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000 -FGAPOP=6.000  
-DELEXT=7.000 -YGAPOP=60.000 -XGAPEXT=60.000 -DELOP=6.000  
-LOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=45  
-MODE=LOCAL -OFTM=pfs -NORM=ext -MINLEN=0 -MAXLEN=200000000  
-USER=US09332522.ecgn1.1.2977 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY  
-WAIT -THREADS=1
```

Search information block:

Query: US-09-332-522B-2

Query length: 1113

Database: EST:*

Database sequences: 5247842

Database length: -2090053206

Search time (sec): 1353.190000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd	Orig	Zscore	EScore	Len	Documentation
gb_est17:D35004	+	110.00	2065.94	3.0e-106	360	D35004 CEIK015H2F yuji kohara u
gb_est17:D68883	+	88.00	1666.92	6.6e-83	360	D68883 CEIK055DZF yuji kohara u
gb_est17:D32385	-	79.00	1360.44	6.0e-67	402	D32385 CEIK012GZR yuji kohara u
gb_est17:D64806	-	58.00	1076.16	4.1e-51	329	D64806 CEIK012GZR yuji kohara u
gb_est17:D67853	+	53.00	980.29	9.0e-46	360	D67853 CEIK087G4F yuji kohara u
gb_est17:D34815	+	48.00	885.06	1.8e-40	360	D34815 CEIK012H2F yuji kohara u
gb_est17:D64735	+	38.00	695.26	6.7e-30	328	D64735 CEIK087G4F yuji kohara u
gb_est17:C63905	+	36.00	656.50	9.7e-28	360	C63905 C63905 yuji kohara unpub
gb_est16:C11998	+	21.00	370.46	8.3e-12	378	C11998 C11998 yuji kohara unpub
gb_est16:C10363	+	16.00	278.11	2.2e-06	252	C10363 C10363 yuji kohara unpub
gb_est16:AM264889	+	10.00	163.23	1.9e-20	274	AM264889 xq36h02.x1 NCI CGAP l
gb_est16:AM719912	+	10.00	158.65	5.09	507	AM719912 LjNEST11h4r lotus japp
gb_est19:AM328320	-	9.00	149.65	16.59	127	AM328320 a2f06ne.fl Neurospora
gb_est14:AM564579	-	9.00	144.24	33.20	272	AM564579 LGL1296.F07.D1.A002.LA
gb_est14:AM748191	-	9.00	143.21	33.31	273	AM748191 sb49c10.y1 Gm.C1011.G1
gb_gss12:AM845611	+	9.00	143.29	36.52	302	AM845611 LMAJ9V1.Lms1e01.y1 Lei
gb_est12:AM065814	+	9.00	143.29	37.50	311	AM065814 AM065814 Rice green sh
gb_gss18:FM0034128	+	9.00	143.04	38.71	322	AM030495 Fugu rubripes GSS seq
gb_est9:AM590282	-	9.00	142.53	41.33	346	AM590282 vml9c03.r1 Knowles SO
gb_est49:H95356	-	9.00	141.87	45.01	380	H95356 yu21f04.s1 Soares fetal
gb_est16:AM711176	+	9.00	141.43	47.59	404	AM711176 es9c1ne.fl Neurospora
gb_est17:AA443338	+	9.00	141.38	47.91	407	AA443338 aa14p10.r1 Soares fetal
gb_est13:Y96709	+	9.00	141.36	48.02	408	Y96709 yesic01.r1 Soares fetal
gb_est17:AM1190367	+	9.00	141.33	48.24	410	AM1190367 q773707.x1 Soares fetal
gb_est19:AM369678	+	9.00	141.07	49.84	425	AM369678 q771606.x1 NCI CGAP B
gb_est14:AM364984	-	9.00	140.81	51.55	441	AM364984 PM4-DT0057-201299-0027
gb_gss12:AM5448702	-	9.00	140.85	52.61	451	AM5448702 RPCI-11-434H18.TV RPCI
gb_est12:AM465887	-	9.00	140.31	54.94	473	AM465887 BP230020B20G4 Soares f
gb_est14:AM429000	+	9.00	140.15	56.10	484	AM429000 BP230020B20G4 Soares f
gb_est14:AM46480	+	9.00	140.10	56.42	487	AM46480 BP230020B20G4 Soares f
gb_gss1:AM036517	+	9.00	139.60	60.21	523	AM036517 f160802.x1 Sugino Kava
gb_est17:AM078348	-	9.00	139.57	60.21	523	AM078348 CTR-RSP-233K2.TF CTR
gb_est18:AM13837	+	9.00	139.47	61.15	541	AM13837 f48c11.y1 Zebrafish w
gb_est19:AM1883741	+	9.00	139.36	62.09	532	AM1883741 DKEF276C2110.T1 762 C
gb_gss18:FM0033875	-	9.00	138.95	65.42	573	AM030495 Fugu rubripes GSS seq
gb_est19:AM1884793	-	9.00	138.85	66.26	581	AM1884793 Fugu rubripes GSS seq
gb_gss13:AM880244	-	9.00	138.78	66.68	587	AM880244 F171009.y1 Zebrafish w
gb_est14:AM583684	+	9.00	138.54	68.95	607	AM583684 HS-5043.B1.B10.T7 RPCI

gb_est25:AM814683	-	9.00	137.88	75.03	666	AM814683 wk66b09.x1 NCI CGAP
gb_gss12:AM086173	-	9.00	136.77	86.53	779	AM086173 nbe0027E12f CUGI R
gb_est20:AM145613	+	8.00	132.44	150.77	98	AM145613 mc59603.x1 Soares mo
gb_est14:AM947344	-	8.00	130.22	200.47	134	AM947344 oq49h11.s1 NCI CGAP
gb_est14:AM92534	-	8.00	130.11	203.19	136	AM92534 ot66h03.s1 Soares t
gb_est24:AM690336	-	8.00	130.11	203.19	136	AM690336 tz15h12.x1 NCI CGAP

seq_name: gb_est17:D35004

seq_documentation_block:

LOCUS	D35004	360 bp	mRNA	EST	08-AUG-1994
DEFINITION	CEIK015H2F yuji kohara unpublished cdna	Caenorhabditis elegans	CDNA		
ACCESSION	D35004				
VERSION	D35004.1	GI:526520			
KEYWORDS	EST.				
SOURCE	Caenorhabditis elegans.				
ORGANISM	Caenorhabditis elegans.				
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.				
AUTHORS	Kohara,Y., Mitsuk,H., Nishigaki,A., Mochizuki,T., Sugimoto,A. and Tabara,H.				
TITLE	Toward an expression map of the C.elegans genome				
JOURNAL	Unpublished (1994)				
COMMENT	On Apr 7, 1998 this sequence version replaced gi:3036136.				
Contact:	yuji kohara				
Gene Library Lab					
National Institute of Genetics					
Yata 1111, Mishima, Shizuoka 411, Japan					
Tel: 81-559-81-6854					
Fax: 81-559-81-6855					
Email: ykohara@lab.nig.ac.jp.					

FEATURES

source

1..360
/organism="Caenorhabditis elegans"
/strain="CB1489 hlm-8(e1489)"
/db_xref="taxon:6239"

/clone="YK15h12"

/clone_lib="yuji kohara unpublished cdna"

/sex="hermaphrodite, male"

/tissue="type='whole animal'"

/dev_stage="varied"

BASE COUNT 123 a 63 c 100 g 73 t 1 others

ORIGIN

alignment_scores:

Quality: 110.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

US-09-332-522B-2 x D35004

Align seg 1/1 to: D35004 from: 1 to: 360

260 PROTRHSLSLVLAVALGILPGLINSEPRIOYSSSRPPOGLINASMETLY 276

29 CCNACTCATTTGGTGAACCAACCAAGTCCAAACCCGAGCAATATGAA 78

276 SGLIULIULVLAVALGILPGLINSEPRIOYSSSRPPOGLINASMETLY 293

79 AGAGGAGCTTCTCGTACCTAGTAAATGATGTCGAGTGAAGTTGAAAC 128

293 TGLVLAVALSLYSSSGLYALACYSSERIALATHASNGLYPROSER 309

129 GGTAAAGAAATTAATAACAGAGCATGTCAGCAGCAATGCGGCATCG 178

310 ARGSESLYGLVLAVALALALYSLVALLIENGILGILVTRALAGLU 326

179 AGGAGTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 228


```

1050 ArgLeuPheValTyrGluLeuThrCysArgMetLeuAnsglyAlaAspR 1066
|||||
224 AGACATATTGCTCTAGAGCTGACGTCGCAATGCTCAACGAGCCACCC 175
1066 OGlnAlaThrTrpSerGly 1072
|||||
174 ACAAGCCACGTCGTCAGGC 156

seq_name: gb_est47:D64806

seq_documentation_block:
LOCUS D64806 329 bp mRNA EST 13-DEC-1995
DEFINITION CEK012GZR Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA
ACCESSION D64806
VERSION D64806.1 GI:1115082
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
AUTHORS Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 329)
Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and
Tabara, H.
Toward an expression map of the C. elegans genome
JOURNAL Unpublished (1994)
COMMENT On Jun 22, 1998 this sequence version replaced gi:3246664.
Contact: Yuj1 Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES
source
1..329
location/Qualifiers
/organism="Caenorhabditis elegans"
/strain="CB1489 hlm-8(e1489)"
/db_xref="taxon:6239"
/clone="YK12g12"
/clone.lib="Yuj1 Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT 81 a 76 c 76 g 95 t 1 others
ORIGIN

alignment_scores:
Quality: 58.00 Length: 58
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x D64806/rev ..
Align seg 1/1 to reverse of: D64806 from: 1 to: 329

1015 ProTyrTrpThrGlnSerPheGlyGlnSerThrPheSerThrLeuTy 1031
|||||
329 CCTTACTGACACAAAGCTTCAAGGACAAATCCACGTTTAGACCTTTA 280
1031 TGlnGluAlaTyrAsnHisTyrAlaIleIleAnsglyThrArgGlyAspC 1048
|||||
279 TCAAGAGCGCATATATCATATGCGATATATATGCGACAAAGGAGATT 230
1048 ySTPARGLeuPheValTyrGluLeuThrCysArgMetLeuAnsglyAla 1064
|||||
229 GTTGAGAGACTATTGCTCTAGAGCTGACGTCGCAATGCTCAACGAGCC 180
1065 AsnProGlnAlaThrTrpSerGly 1072
|||||

```

```

179 AACCCACAGCCACGTCGTCAGGC 156

seq_name: gb_est47:D67853

seq_documentation_block:
LOCUS D67853 360 bp mRNA EST 07-DEC-1995
DEFINITION CEK087G4F Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA
ACCESSION D67853
VERSION D67853.1 GI:1105632
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
AUTHORS Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 360)
Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and
Tabara, H.
Toward an expression map of the C. elegans genome
JOURNAL Unpublished (1994)
COMMENT Contact: Yuj1 Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES
source
1..360
location/Qualifiers
/organism="Caenorhabditis elegans"
/strain="CB1489 hlm-8(e1489)"
/db_xref="taxon:6239"
/clone="YK87g4"
/clone.lib="Yuj1 Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT 98 a 86 c 85 g 88 t 3 others
ORIGIN

alignment_scores:
Quality: 53.00 Length: 53
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x D67853 ..
Align seg 1/1 to: D67853 from: 1 to: 360

428 LeuProTyrProGluProIleGlnTyrThrGluTyrSerAlaArgSerPr 444
|||||
3 TTACCGTACCCGGAACCAATTCATACCTGATGATCTGCCGATCACCC 52
444 OValGluSerSerProSerProProArgAsnGluArgGlyArgSerArgm 461
|||||
53 CGTGAATCATCTCCTCTCCACCTAGAAATGAGAGAAAGATCACGAA 102
461 eTserThrThrThrProMetLysAnsglyThrArgAspGlySerSerLys 477
|||||
103 TGAGCACAACACACTCCTATGAAGATGAGAACTAGAGATGATCTTCGAA 152
478 ValThrLeu 480
|||||
153 GTTACCTTN 161

seq_name: gb_est47:D34815

seq_documentation_block:
LOCUS D34815 360 bp mRNA EST 08-AUG-1994
DEFINITION CEK012HZF Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone YK12h12 5', mRNA sequence.

```

ACCESSION D34815
 VERSION D34815.1 GI:526331
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 360)
 AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.
 TITLE Toward an expression map of the C.elegans genome
 JOURNAL Unpublished (1994)
 COMMENT On Apr 7, 1998 this sequence version replaced gi:3035947.
 CONTACT: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855
 Email: ykohara@lab.nig.ac.jp.
 FEATURES
 source
 Location/Qualifiers
 1..360
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(e1489)"
 /db_xref="taxon:6239"
 /clone="YK12h12"
 /clone_lib="Yuji Kohara unpublished cDNA"
 /sex="hermaphrodite, male"
 /tissue_type="whole animal"
 /dev_stage="varied"
 BASE COUNT 119 a 106 c 58 g 74 t 3 others
 ORIGIN
 alignment_scores:
 Quality: 48.00 Length: 48
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-332-522B-2 x D34815 ..
 Align seg 1/1 to: D34815 from: 1 to: 360
 200 HisTleasnGLYTrMetSerProTYrAspGlnAlaGlnGLYProSerGI 216
 |||||||
 77 CATATCAACGGGTACATGCTCCATACGACCACCTCAAGNCATCAGG 126
 216 YProSerTYrTYrSerGlnHisGlnSerProProProHisHisHisH 233
 |||||||
 127 ACCATCATATTACTCACACACCATCATCTCCACCACTCATATCACACC 176
 233 tshHisHisPrometProGLysIleHisGlnAsnProGlnGlnVal 247
 |||||||
 177 ATCAACCCACCGATGCCAAAATCATCATGAGAACCTGAACAAGTG 220
 seq_name: gb_est47:D64735
 seq_documentation_block:
 LOCUS D64735 328 bp mRNA EST 13-DEC-1995
 DEFINITION CELR087G4R Yuji Kohara unpublished cDNA Caenorhabditis elegans CDNA
 clone YK87g4 3', mRNA sequence.
 ACCESSION D64735
 VERSION D64735.1 GI:1116421
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 328)
 AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.
 TITLE Toward an expression map of the C.elegans genome

JOURNAL Unpublished (1994)
 COMMENT On Oct 6, 1998 this sequence version replaced gi:3705561.
 CONTACT: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855
 Email: ykohara@lab.nig.ac.jp.
 FEATURES
 source
 Location/Qualifiers
 1..328
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(e1489)"
 /db_xref="taxon:6239"
 /clone="YK87g4"
 /clone_lib="Yuji Kohara unpublished cDNA"
 /sex="hermaphrodite, male"
 /tissue_type="whole animal"
 /dev_stage="varied"
 BASE COUNT 81 a 75 c 75 g 91 t 6 others
 ORIGIN
 alignment_scores:
 Quality: 38.00 Length: 38
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-332-522B-2 x D64735/rev ..
 Align seg 1/1 to reverse of: D64735 from: 1 to: 328
 1016 TYrTPThrGlnSerPheLysGLysGlnSerThrPheSerThrLeuTYrGI 1032
 |||||||
 327 TACTGACACAAAGCTTCAAGGNCATCCACGTTAGTACGCTTATCA 278
 1032 nGuaLaIeTYrAsnHisTYrAlaLeIleLysnGLYThrArgGLYAspCysT 1049
 |||||||
 277 AGAGCGGTATTAATCATATGCGATTATATGAGAACGGAGATTTGT 228
 1049 rPArGLeupPheVal 1053
 |||||||
 227 GGAGACTATTTC 214
 seq_name: gb_est47:C63905
 seq_documentation_block:
 LOCUS C63905 360 bp mRNA EST 22-SEP-1997
 DEFINITION C63905 Yuji Kohara unpublished cDNA Caenorhabditis elegans CDNA
 clone YK349h3 5', mRNA sequence.
 ACCESSION C63905
 VERSION C63905.1 GI:2422610
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 360)
 AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A. and Nishigaki,A.
 TITLE Expression map of the C.elegans genome
 JOURNAL Unpublished (1996)
 COMMENT On Apr 7, 1998 this sequence version replaced gi:3035983.
 CONTACT: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855
 Email: ykohara@lab.nig.ac.jp.
 FEATURES
 source
 Location/Qualifiers
 1..360

```

/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="YK349h3"
/clone_lib="Yuj1 Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT      121 a      92 c      73 g      66 t      8 others
ORIGIN

alignment_scores:
  Quality:      36.00      Length:      36
  Ratio:        1.000      Gaps:      0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522b-2 x C63905

Align seg 1/1 to: C63905 from: 1 to: 360

276 LysGluGluLeuLeuArgLeuValAsnMetSerProSerGluValG1 292
|||||
209 AAGAGAGACTTCTGCTGCTACTAGTACATCTCTCCAGATGAACTTGA 258
|||||
292 uArgLeuLysAsnLysLysSerGlyAlaCysSerAlaThrAsnGlyProS 309
|||||
259 ACGGTTAAAGAAATAAATAATCNGAGCAGATGTCAGCAGCAATGGCCAT 308
|||||
309 erArgSer 311
|||||
309 CGAGGAGT 316

seq_name: gb_est46:C11998

seq_documentation_block:
LOCUS      C11998      378 bp      mRNA      EST      28-DEC-1998
DEFINITION C11998 Yuj1 Kohara unpublished cDNA Caenorhabditis elegans CDNA
clone YK144h7 5', mRNA sequence.
ACCESSION  C11998
VERSION    C11998.1 GI:1559551
KEYWORDS
SOURCE
ORGANISM
REFERENCE
  1 (bases 1 to 378)
  Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A.,
  Sano, M., Miyata, A. and Nishigaki, A.
  Expression map of the C.elegans genome
  Unpublished (1996)
  On Nov 2, 1998 this sequence version replaced gi:3828838.
  Contact: Yuj1 Kohara
  Gene Library Lab
  National Institute of Genetics
  Yata 1111, Mishima, Shizuoka 411, Japan
  Tel: 81-559-81-6854
  Fax: 81-559-81-6855
  Email: ykohara@lab.nig.ac.jp.
  Location/Qualifiers
    1..378
    /organism="Caenorhabditis elegans"
    /strain="CB1489 him-8(e1489)"
    /db_xref="taxon:6239"
    /clone="YK144h7"
    /clone_lib="Yuj1 Kohara unpublished cDNA"
    /sex="hermaphrodite, male"
    /tissue_type="whole animal"
    /dev_stage="varied"
BASE COUNT      93 a      96 c      90 g      89 t      10 others
ORIGIN

```

```

alignment_scores:
  Quality:      21.00      Length:      21
  Ratio:        1.000      Gaps:      0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522b-2 x C11998

Align seg 1/1 to: C11998 from: 1 to: 378

979 AlaSerGlyValSerGlyIleGlnGlyThrArgArgMetAlaTy 995
|||||
203 GCATCGGTGTGTCTGGAATTCAGGAAGGTACACGCCGAAATGCCCTA 252
|||||
995 rGluTrpIleMet 999
|||||
253 CGAATGCATTATG 265

seq_name: gb_est46:C10363

seq_documentation_block:
LOCUS      C10363      252 bp      mRNA      EST      28-DEC-1998
DEFINITION C10363 Yuj1 Kohara unpublished cDNA Caenorhabditis elegans CDNA
clone YK144h7 3', mRNA sequence.
ACCESSION  C10363
VERSION    C10363.1 GI:1535434
KEYWORDS
SOURCE
ORGANISM
REFERENCE
  1 (bases 1 to 252)
  Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A.,
  Sano, M., Miyata, A. and Nishigaki, A.
  Expression map of the C.elegans genome
  Unpublished (1996)
  On Jul 8, 1999 this sequence version replaced gi:5422657.
  Contact: Yuj1 Kohara
  Gene Library Lab
  National Institute of Genetics
  Yata 1111, Mishima, Shizuoka 411, Japan
  Tel: 81-559-81-6854
  Fax: 81-559-81-6855
  Email: ykohara@lab.nig.ac.jp.
  Location/Qualifiers
    1..252
    /organism="Caenorhabditis elegans"
    /strain="CB1489 him-8(e1489)"
    /db_xref="taxon:6239"
    /clone="YK144h7"
    /clone_lib="Yuj1 Kohara unpublished cDNA"
    /sex="hermaphrodite, male"
    /tissue_type="whole animal"
    /dev_stage="varied"
BASE COUNT      57 a      63 c      62 g      66 t      4 others
ORIGIN

alignment_scores:
  Quality:      16.00      Length:      16
  Ratio:        1.000      Gaps:      0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522b-2 x C10363/rev

Align seg 1/1 to reverse of: C10363 from: 1 to: 252

1091 ArgArgSerAlaGlnProAspAlaPheHisLeuHisThrLeuVallys 1106
|||||
50 CGAGCTCGGCTCAACCGGACGCAATTCATCTTCACTACACTGTTAA 3

```

```

seq_name: gb_est40:AW264889
seq_documentation_block:
  LOCUS      AW264889      274 bp      mRNA      EST      28-DEC-1999
  DEFINITION  xq36102.x1 NCI_CGAP_Lu28 Homo sapiens cDNA clone IMAGE:2752755 3'
  ACCSSION   AM264889
  VERSION    AM264889.1  GI:6641705
  KEYWORDS   EST.
  SOURCE      human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  REFERENCE  1 (bases 1 to 274)
             NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
             National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
             Unpublished (1997)
  JOURNAL    Contact: Robert Strausberg, Ph.D.
             Tel: (301) 496-1550
             Email: Robert.Strausberg@nih.gov
             Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
             Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life
             Technologies, Inc. cDNA Library Arrayed by: Christina Prange, The
             I.M.A.G.E. Consortium DNA sequencing by: Washington University
             Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/dbp/image/image.html

FEATURES
  source
    1..274
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:2752755"
    /clone_1lb="NCI_CGAP_Lu28"
    /tissue_type="two pooled squamous cell carcinomas"
    /note="Organ: Lung; Vector: PCMV-SPORT6; Site_1: SalI;
    Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
    Library constructed by Life Technologies."
  BASE COUNT      43 a      84 c      91 g      56 t
  ORIGIN
    43 a      84 c      91 g      56 t

alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-09-332-522B-2 x AW264889
  ..

Align seg 1/1 to: AW264889 from: 1 to: 274
  100 AspleuglyglyArgGlyProSerLeu 109
  ||||||||||||||||||||||||||||
  68 GACCTGGGGGGGGCGAGCCGACGCTG 97

seq_name: gb_est46:AW719912
seq_documentation_block:
  LOCUS      AW719912      507 bp      mRNA      EST      19-APR-2000
  DEFINITION  LjNEST114r Lotus japonicus nodule library 5 and 7 week-old Lotus
  japonicus cDNA 5', mRNA sequence.
  ACCSSION   AW719912
  VERSION    AW719912.1  GI:7614436
  KEYWORDS   EST.
  SOURCE      Lotus japonicus.
  ORGANISM   Lotus japonicus

```

```

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales;
Fabaceae; Papilionoideae; Lotus.
1 (bases 1 to 507)
Colabatch,G., Freund,S., Trevasakis,B and Ugdvardi,M.
Lotus japonicus root nodule ESTs: tools for functional genomics
Unpublished (2000)
On Jan 6, 2000 this sequence version replaced gi:6675537.
Contact: Ugdvardi MK
Molecular Plant Nutrition
Max Planck Institute of Molecular Plant Physiology
Am Muehlenberg 1, 14476 Golm, Germany
Fax: 49 331 567 8250
Email: ugdvardi@mpimp-golm.mpg.de
Seq primer: T7
High quality sequence stop: 507.
Location/Qualifiers
  1..507
  /organism="Lotus japonicus"
  /cultivar="Gifu (8-129)"
  /db_xref="taxon:34305"
  /clone_1lb="Lotus japonicus nodule library 5 and 7
  week-old"
  /dev_stage="5 and 7 week-old plants"
  /note="Organ: Nodule; Vector: pSPORT1; Site_1: SalI;
  Site_2: NotI; The library was prepared using mRNA
  extracted from nodules of 5 and 7 week-old Lotus plants.
  Nodules were induced by, and contained Mesorhizobium
  strain R7A."
  BASE COUNT      116 a      194 c      99 g      98 t
  ORIGIN
    116 a      194 c      99 g      98 t

alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-09-332-522B-2 x AW719912
  ..

Align seg 1/1 to: AW719912 from: 1 to: 507
  227 ProProHSHSHSHSHSHSHSHPro 236
  ||||||||||||||||||||||||||||
  23 CCACCACCACCACACACACACACCC 52

seq_name: gb_est19:AI328320
seq_documentation_block:
  LOCUS      AI328320      127 bp      mRNA      EST      28-DEC-1998
  DEFINITION  a2f06ne.f1 Neurospora crassa evening cDNA library Neurospora crassa
  cDNA clone a2f06ne 5', mRNA sequence.
  ACCSSION   AI328320
  VERSION    AI328320.1  GI:4064879
  KEYWORDS   EST.
  SOURCE      Neurospora crassa.
  ORGANISM   Neurospora crassa
             Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae;
             Neurospora.
  REFERENCE  1 (bases 1 to 127)
             Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.
             Two Neurospora crassa EST Databases
             Unpublished (1998)
             Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
             Department of Chemistry and Biochemistry
             Advanced Center for Genome Technology, University of Oklahoma
             620 Parrington Oval, Norman, OK 73019, USA
             Tel: 405 325 4912
             Fax: 405 325 7762
             Email: broe@ou.edu
             We anticipate the future release of the cDNA clones to the Fungal
             Genetics Stock Center

```

Seq primer: Universal Forward Primer
High quality sequence stop: 118.

FEATURES

source

1. 127

/organism="Neurospora crassa"

/strain="Strain 30-7 (bd; A)"

/db_xref="taxon:5141"

/clone_1lb="a2f06e"

/clone_1lb="Neurospora crassa evenning CDNA library"

/tissue_type="tissue harvested following 22hr growth in dark"

/note="Vector: pBluescript SK+; Site.1: XbaI; Site.2: EcoRI; See: Bell-Pedersen, D., et al. PNAS 93:13096, 1996. 5' end of CDNA cloned into XbaI site of pBluescript; 3' end of CDNA cloned into EcoRI site of pBluescript"

BASE COUNT 31 a 19 c 52 g 25 t

ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x A1328320/rev ..

Align seg 1/1 to reverse of: A1328320 from: 1 to: 127

seq_name: gb_est44:AM564579

seq_documentation_block:

LOCUS

AM564579

272 bp mRNA

EST 10-MAR-2000

DEFINITION LG1.296_F07.b1.A002 Light Grown 1 (LG1) Sorghum bicolor CDNA, mRNA

sequence.

ACCESSION AM564579

VERSION AM564579.1 GI:7218457

KEYWORDS EST.

SOURCE

ORGANISM

Sorghum bicolor

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.

1 (bases 1 to 272)

Cordonnier-Pratt, M.M., Gingle, A., Pratt, L. and Peterson, A.

An EST database from Sorghum

Unpublished (1999)

On Jun 22, 1998 this sequence version replaced g1:3246846.

Contact: Cordonnier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: JEN REV

High quality sequence stop: 260

POLYA-No.

Location/Qualifiers

1. 272

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_1lb="Light Grown 1 (LG1)"

/note="Organ: 10- to 14-day-old light-grown (greenhouse) seedlings; Vector: Lambda Zap; Site.1: XhoI; Site.2: EcoRI; The library was made from poly-A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were

BASE COUNT 79 a 43 c 88 g 62 t

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x AM564579/rev ..

Align seg 1/1 to reverse of: AM564579 from: 1 to: 272

227 PROPOPHILSHISHISHISHIS 235
|||||
217 CCCCCCTCATCATCATCATCATCAT 191

seq_name: gb_est24:A1748191

seq_documentation_block:

LOCUS

A1748191

273 bp mRNA

EST 13-DEC-1999

DEFINITION

sb49c10.y1.Gm-cl011.Glycine max CDNA clone GENOME SYSTEMS CLONE ID: Gm-cl011.235 5', mRNA sequence.

ACCESSION

A1748191

VERSION A1748191.1 GI:5126455

KEYWORDS EST.

SOURCE

ORGANISM

Glycine max

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

1 (bases 1 to 273)

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvett, J., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

Public Soybean EST Project

Unpublished (1999)

On Apr 7, 1998 this sequence version replaced g1:3036048.

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com

Putative full length read

vector to vector length is 303

High quality sequence stop: 266.

Location/Qualifiers

1. 273

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl011-235"

/clone_1lb="Gm-cl011"

/tissue_type="Immature cotyledons of greenhouse grown plants"

/lab_host="DH10B"

/note="Vector: pBluescript II SK+; Site.1: EcoRI; Site.2: XhoI; This CDNA library was constructed from mRNA isolated from immature cotyledons (100-200mg) of greenhouse grown plants. The CDNA library was prepared using the Life Technologies superscript CDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly

(dT) sequence with a Not I restrictions site. Sal I linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-Sal I restriction site of the pSPORT 1 vector. The ligated cDNA fragments were transformed into E. coli Electromax DH10B host cells. This library was constructed by Dr. Lila Vodkin and Dr. Anu Khanna."

BASE COUNT 100 a 40 c 65 g 68 t
ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522b-2 x A1748191/rev ..

Align seg 1/1 to reverse of: A1748191 from: 1 to: 273

969 SerLeuArgLeuPheSerThrSer 977
|||||
184 AGCTTAGACGTCTCAGACCTCCTCA 158

seq_name: gb_gss12:A0849611

seq_documentation_block:

LOCUS A0849611 302 bp DNA GSS 15-OCT-1999
DEFINITION LMAJFV1_lm51e01.v1 Leishmania major FV1 random genomic library
survey sequence.
ACCESSION A0849611

VERSION A0849611.1 GI:6054259

KEYWORDS GSS.
SOURCE Leishmania major.
ORGANISM Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.

REFERENCE 1 (bases 1 to 302)

Akopyants, N.S., Beverley, S.M., Clifton, S., Marra, M., Hillier, L.,
Chinwalla, A., Pape, D., Martin, J., Wylie, T., Blistain, A.,
Schmitt, A., Person, B., Theising, B., Ritter, E., Ronko, T.,
Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M.,
Harvey, N., McCann, R., Tsagarashvili, R., Williams, T., Jackson, Y.,
Bowers, Y., Swaller, T., Waterston, R. and Wilson, R.
WashU Leishmania Project
Unpublished (1999)

TITLE
JOURNAL
COMMENT
Contact: Akopyants, NS / Beverley, SM
WashU Leishmania Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu

Library construction: Natalia S. Akopyants, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
if using this information please cite:
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
Friedlin strain VI genome by shotgun sequencing' and the Washington
University Genome Sequencing Center for information on obtaining
clone material. Please contact: Natalia S. Akopyants Ph.D.
(natalia@porcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
(beverley@porcim.wustl.edu)

Seq primer: -40Rp from Gibco
Class: shotgun
High quality sequence stop: 281.

FEATURES
source
Location/Qualifiers
1..302
/organism="Leishmania major"
/strain="Friedlin strain VI"

/db_xref="taxon:5664"
/clone="LMAJFV1_lm51e01"
/clone_lib="Leishmania major FV1 random genomic library"
/lab_host="TOP10 (Invitrogen)"
/note="Vector: pZero-2 (Invitrogen); Site_1: EcoRV;
genomic DNA was isolated from stationary phase cells. For
this library, DNA was sheared to give a tight size
distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
polymerase, dephosphorylated with Shrimp Alkaline
Phosphatase and ligated into pZero-2 vector's EcoRV site."

BASE COUNT 74 a 130 c 43 g 55 t
ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522b-2 x A0849611 ..

Align seg 1/1 to: A0849611 from: 1 to: 302

227 ProProHisHisHisHisHisHis 235
|||||
193 CCTCTCTCTCCACCCACCCACAC 219

seq_name: gb_est129:A0065814

seq_documentation_block:

LOCUS A0065814 311 bp mRNA EST 02-JUN-1999
DEFINITION A0065814 Rice green shoot Oryza sativa cDNA clone S1173_1A, mRNA
sequence.
ACCESSION A0065814
VERSION A0065814.1 GI:4968910
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.

REFERENCE 1 (bases 1 to 311)

Rice cDNA from green shoot
Unpublished (1996)
On Apr 7, 1998 this sequence version replaced gi.3035543.
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@nri.affrc.go.jp
PROJECT = RGP

TITLE
JOURNAL
COMMENT
Location/Qualifiers
1..311
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="S1173_1A"
/clone_lib="Rice green shoot"
/note="Green shoot (8 days old)"

BASE COUNT 120 a 57 c 84 g 48 t 2 others
ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x AU065814/rev ..

Align seg 1/1 to reverse of: AU065814 from: 1 to: 311

227 ProProProHSHSHSHSHSHSHS 235
 |||||||
 35 CCTCCTCCTCATCATCATCATCAT 9

seq_name: gb_gss18:FR0034128

seq_documentation_block:
 LOCUS FR0034128 322 bp DNA GSS 27-JUN-1998
 DEFINITION Fugu rubripes GSS sequence, clone 189A09ac3, genomic survey
 sequence.
 ACCESSION AL030495
 VERSION AL030495.1 GI:3272609
 KEYWORDS GSS: genome survey sequence.
 SOURCE Fugu rubripes.
 ORGANISM Fugu rubripes.
 Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
 Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
 Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.
 1 (bases 1 to 322)
 Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umranta,Y.,
 Williams,G. and Brenner,S.
 DIRECT Submission
 Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource
 Centre, Hinxton, Cambridge, CB10 1SB, UK. Email:
 biohelp@hmp.mrc.ac.uk
 Vector: pBluescript II KS
 V-type: phagemid
 PRIMER: KS
 DESC: One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

FEATURES
 source Location/Qualifiers
 1..322
 /organism="Fugu rubripes"
 /db_xref="taxon:31033"
 /clone_lib="cosmid 189A09"
 /clone="189A09ac3"

BASE COUNT 90 a 88 c 62 g 69 t 13 others

ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x FR0034128 ..

Align seg 1/1 to: FR0034128 from: 1 to: 322

227 ProProProHSHSHSHSHSHSHS 235
 |||||||
 113 CCTCCTCCTCATCATCATCATCAT 139

seq_name: gb_est9:AA590282

seq_documentation_block:
 LOCUS AA590282 346 bp mRNA EST 16-SEP-1997
 DEFINITION vm19q03.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
 clone IMAGE:990676 5', mRNA sequence.
 ACCESSION AA590282
 VERSION AA590282.1 GI:2403662
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 346)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE The Mashu-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2151790.
 Contact: Maria W/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:562956
 Putative full length read
 vector to vector length is 347.

FEATURES
 source Location/Qualifiers
 1..346
 /organism="Mus musculus"
 /strain="B6D2 F1/J"
 /db_xref="taxon:10090"
 /clone_image="990676"
 /clone_lib="Knowles Solter mouse blastocyst B1"
 /tissue_type="blastocyst"
 /dev_stage="embryo (pre-implantation)"
 /lab_host="DH10B"
 /note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally from mRNA prepared
 from 800 blastocysts. Primer: SalI(dT):
 5'-CGGTGACCGTCGACCGTGTGTGTGT-3'. cDNAs were
 cloned into the NotI/SalI sites of a pSPORT vector (Life
 Technologies) and B3." Two different size selections: B1 (larger
 inserts) and B3."

BASE COUNT 86 a 84 c 86 g 90 t

ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522B-2 x AA590282/rev ..

Align seg 1/1 to reverse of: AA590282 from: 1 to: 346

128 GlnThrSerProPserGlyGlyTyr 136
 |||||||
 134 CAGACTTCTCCGACCTCGTGATAC 108

seq_name: gb_est49:H95356

seq_documentation_block:
 LOCUS H95356 380 bp mRNA EST 25-NOV-1996
 DEFINITION YU21F04.s1 Soares fetal liver spleen INFUS Homo sapiens cDNA clone
 IMAGE:234463 3', mRNA sequence.
 ACCESSION H95356
 VERSION H95356.1 GI:1102989
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
 REFERENCE 1 (bases 1 to 380)
 AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
 Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
 Parsons,J., Rifein,L., Rohlfing,T., Soares,M., Tan,F.,


```

/c/clone_1lb="Soares.NhMHPu_S1"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalised libraries (melanocyte 2NDHw, pregnant uterus
NbhpU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of T.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT      89 a      107 c      143 g      68 t
ORIGIN

Alignment_scores:
Quality:          9.00      Length:          9
Ratio:            1.000     Gaps:           0
Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
US-09-532-522B-2 x AA443238 ..

Align seg 1/1 to: AA443238 from: 1 to: 407

975 ThrseserArgAlaserGlyValVal 983
|||||
279 ACCTCCTCCAGAGCCTCAGAGTGTC 305

seq_name: gb_est53:T96709

seq_documentation_block:
LOCUS       T96709             408 bp      mRNA      EST      27-MAR-1995
DEFINITION  y551c01.r1 Soares fetal liver spleen INFILS Homo sapiens CDNA clone
IMAGE:121248 5' , mRNA sequence.
ACCESSION   T96709
VERSION     T96709.1 GI:735333
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 408)
AUTHORS    Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hulman,M., Kucada,T., Le,M., Lennon,G., Merit,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevisakis,E., Waterston,R., Williamson,A., Woldmann,P. and
Wilson,R.
JOURNAL     The WashU-Merk EST Project
COMMENT     Unpublished (1995)
On Apr 14, 1993 this sequence version replaced gi:693317.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1009
High quality sequence stops: 341 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1009 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 341.
Location/Qualifiers
1. 408
/organism="Homo sapiens"
/db_xref="GDB:473793"

/c/clone_1lb="Soares.NhMHPu_S1"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalised libraries (melanocyte 2NDHw, pregnant uterus
NbhpU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of T.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT      92 a      119 c      122 g      65 t
ORIGIN

Alignment_scores:
Quality:          9.00      Length:          9
Ratio:            1.000     Gaps:           0
Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
US-09-532-522B-2 x T96709 ..

Align seg 1/1 to: T96709 from: 1 to: 408

975 ThrseserArgAlaserGlyValVal 983
|||||
77 ACCTCCTCCAGAGCCTCAGAGTGTC 103

seq_name: gb_est17:A1190367

seq_documentation_block:
LOCUS       A1190367             410 bp      mRNA      EST      28-OCT-1998
DEFINITION  qd37n07.x1 Soares_fetal_heart_NBH19W Homo sapiens CDNA clone
IMAGE:1731709 3' similar to SW:CIQA_HUMAN P02745 COMPLEMENT C1Q
SUBCOMPONENT, A CHAIN PRECURSOR. ; , mRNA sequence.
ACCESSION   A1190367
VERSION     A1190367.1 GI:3741576
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 410)
AUTHORS    NCI-GAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Jun 15, 1998 this sequence version replaced gi:3223706.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1350
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 518 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
1. 410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="IMAGE:1731709"
/clone_1lb="Soares_fetal_heart_NBH19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: liver and Spleen; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Pac I; Site 2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACGTGAGAGAATTAATTAAAGATCTTTTCTTTTTTTTTTTT 3'] ,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia). digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

```


Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment_block:

US-09-332-522B-2 x AW364984/rev ..

Align seg 1/1 to reverse of: AW364984 from: 1 to: 441

608 AsparagserleuproserProglyval 616

|||||

seq_name: gb_gss8:AO552484

seq_documentation_block:

LOCUS AO552484 451 bp DNA GSS 28-MAY-1999

DEFINITION RPCI-11-434H18.TV RPCI-11 Homo sapiens genomic clone RPCI-11-434H18, genomic survey sequence.

ACCESSION AO552484

VERSION AO552484.1 GI:4911661

KEYWORDS GSS.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 451)

AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building

JOURNAL Unpublished (1997)

COMMENT On Dec 13, 1999 this sequence version replaced gi:4211931.

Other GSSs: RPCI-11-434H18.TJ

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong

(pieter@ejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genet. cs (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.

Seq primer: T7

Class: BAC ends

Location/Qualifiers

1. 451

/organism="Homo sapiens"

/db_xref="GDB:7666457"

/db_xref="taxon:9606"

/clone="RPCI-11-434H18"

/clone_id="RPCI-11"

/sex="Male"

/cell_type="Lymphocytes"

/note="Vector: pBAC3.6, Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC library"

BASE COUNT 74 a 83 c 77 g 217 t

Alignment_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

Alignment_block:

US-09-332-522B-2 x AO552484/rev ..

Align seg 1/1 to reverse of: AO552484 from: 1 to: 451

419 MetargylsThnleuGlnAsnArg 427

|||||

281 ATGAGAGAGCCCTCCATCTCAGTGTT 23

seq_name: gb_est42:AM444702

seq_documentation_block:

LOCUS AM444702 473 bp mRNA EST 17-FEB-2000

DEFINITION UI-R-B13-a1x-a-05-0-UI.s1 NCI_CGAP_Sub5 Homo sapiens cDNA clone IMAGE:2733201 3', mRNA sequence.

ACCESSION AM444702

VERSION AM444702.1 GI:6986464

KEYWORDS EST.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 473)

AUTHORS National Cancer Institute, Cancer genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. cDNA library preparation: M.B. Soares Lab Clone distribution:

NCI-CGAP clone distribution information can be found through the

I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/dbp/image/image.html The following repetitive

elements were found in this cDNA sequence: 1-42,

>AT-richlow complexity 126-151, >AT-richlow complexity

Seq primer: M13 forward

Location/Qualifiers

1. 473

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2733201"

/clone_id="NCI_CGAP_Sub5"

/lab_host="DH10B (Life Technologies)"

/note="Vector: p773D-Pac (Pharmacia) with a modified

polylinker. Site_1: Not I; Site_2: Eco RI; NCI CGAP Subs

is a subtracted library derived from NCI_CGAP_Sub4. The

NCI_CGAP_Sub5 library had 3 million recombinants. A

single-stranded DNA preparation of NCI_CGAP_Sub4 was used

as a tracer in a subtractive hybridization with a driver

comprising: the IMAGE pool (NCI_CGAP_K1d3 pool 1 LLAM

3334-3337, 3682-3683, 3798-3803 (IMAGE Clonoids

1323276-1323911, 1456008-1456775, 1500552-1502855).

NCI_CGAP_K1d5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778

(IMAGE Clonoids 1323912-1325831, 1471368-1472903,

1492104-1493255); NCI_CGAP_Ln5 pool 1 LLAM 3575-3582,

3851-3854 (IMAGE Clonoids

1414920-1417991, 1520904-1522439); NCI CGAP GC4 pool 1 LLAM

3164-3167, 3716-3720, 3733-3735 (IMAGE Clonoids

1257096-1258631, 1469064-1470983, 1475592-1476743);

NCI_CGAP_Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068

(IMAGE Clonoids 985608-986759, 1101192-1101959,

1217928-1220615); NCI_CGAP_Co10 pool 1 LLAM 2644-2653,

1057416-1061255, 1144584-1145351) (10% of the driver

population), plus a pool of 3,840 arrayed clones from

NCI_CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and

NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (10% of

the driver population), plus a pool of 11,136 clones from

NCI CGAP Subs (IMAGE Clonoids 2712456-2723591) (10% of the

driver population) plus a pool of 5,472 clones from

NCI_CGAP_Sub4 (IMAGE Clonoids 2723592-2728569) (70% of the

driver population). Subtraction was performed as

previously described [Bonaldo, Lennon & Soares (1996):
 Normalization and Subtraction: Two Approaches To
 Facilitate Gene Discovery. Genome Research 6, 791-806.
 TAG-LIB-MCL-CGAP-Col0; TAG-TISSUE-colon; TAG-SEQ-AAACG"]
 BASE COUNT 96 a 120 c 127 g 130 t
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522b-2 x AW444702/rev ..

Align seg 1/1 to reverse of: AW444702 from: 1 to: 473

975 ThrSerSerArgAlaSerGlyValVal 983
 |||||
 350 ACCTCCTCCAGAGCCTCAGAGTGTGTC 324

seq_name: gb_est42:AW465887

seq_documentation_block: 484 bp mRNA EST 24-FEB-2000
 LOCUS AW465887
 DEFINITION BP230020B20G4 Soares normalized bovine placenta Bos taurus cDNA
 clone BP230020B20G4 5', mRNA sequence.
 ACCESSION AW465887
 VERSION AW465887.1 GI:7036055
 KEYWORDS EST.
 SOURCE Bos taurus.
 ORGANISM Bos taurus.
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 484)
 Lewin,H.A., Soares,M.B., Rebelz,M., Pardinas,J., Liu,L. and
 Larson,J.H.
 TITLE Bosvine ESTs
 JOURNAL Unpublished (2000)
 COMMENT Contact: Lewin, H. A.
 W. M. Keck Center for Comparative and Functional Genomics
 University of Illinois at Urbana-Champaign
 340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
 61801, USA
 Tel: 217 333 5998
 Fax: 217 244 5617
 Email: h-lewin@uiuc.edu

Funding for cattle EST sequencing was provided by the USDA National
 Research Initiative, Animal Genome Resource Grant AG 99-3205-8534
 to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
 from Washington University Genome Center. Vector Trim1 g:
 Cross-match from Washington University Genome Center PHRAP suite.
 Sequences submitted are vector free and at least 200 bp in length.
 PCR Primers
 FORWARD: TATACGACTCCTATAGG
 BACKWARD: ATTACCTCTACTAAG
 Insert Length: 484 Std Error: 0.00
 Plate: BP230020B20 row: G column: 4
 Seq primer: AGCGATACATTCATCCACAGCA
 High quality sequence stop: 484.
 Location/Qualifiers
 1..484

FEATURES

source
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone="BP230020B20G4"
 /clone_lib="Soares normalized bovine placenta"
 /sex="female"
 /lab_host="DH10B"
 /note="Organ: placenta; Vector: pT7T3pac; Site:1: EcoRI;
 Site:2: NotI; The cDNA library was contributed by the
 Soares laboratory and it was constructed and normalized

as described by Bonaldo, M.F., Lennon, G. and Soares,
 M.B. (1996), Genome Research 6(9): 791-806.
 BASE COUNT 129 a 132 c 129 g 90 t 4 others
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522b-2 x AW465887 ..

Align seg 1/1 to: AW465887 from: 1 to: 484

227 PROProHSHSHSHSHSHSHSHS 235
 |||||
 285 CCACACACACACACACACACACAT 311

seq_name: gb_est41:AW329000

seq_documentation_block: 487 bp mRNA EST 28-JAN-2000
 LOCUS AW329000
 DEFINITION N200194e rootphos(-) Medicago truncatula cDNA clone MHRP-15B02,
 mRNA sequence.
 ACCESSION AW329000
 VERSION AW329000.1 GI:6799595
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales;
 Fabaceae; Papilionoideae; Medicago.
 1 (bases 1 to 487)
 Harrison,M.J., Liu,J., Gonzales,M.B. and Ellis,L.
 TITLE ESTs from phosphate starved roots
 JOURNAL Unpublished (1999)
 COMMENT On Jul 7, 1999 this sequence version replaced gi:5866082.
 Contact: Maria J. Harrison
 Plant Biology Division
 The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73401, USA
 Tel: 580-223-5810
 Fax: 580-221-7380
 Email: mjharrison@noble.org

Developmental stage: phosphate starved; Date: 01/19/00; Submitted
 to the database of Expressed Sequence Tags (dbEST) on 01/26/00;
 More information is available at
 'http://chrstyle.tamu.edu/medicago'.
 Seq primer: T3.
 Location/Qualifiers
 1..487

FEATURES

source
 /organism="Medicago truncatula"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone="MHRP-15B02"
 /clone_lib="rootphos(-)"
 /tissue_type="roots"
 /note="Vector: pBluescript SK-; Site:1: EcoRI; Site:2:
 XhoI; cDNA was prepared from polyA+ enriched RNA. The
 cDNA was directionally ligated into the Unizap XR vector
 from Stratagene and packaged using Gigapack III Gold
 packaging extracts. Plasmids containing cDNA inserts
 were excised from the recombinant lambda-zap phage using
 Ex-assist helper phage and propagated in SOUR cells."

BASE COUNT 107 a 143 c 96 g 141 t
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x AW329000 ..

Align seg 1/1 to: AW329000 from: 1 to: 487

227 PROPROHSHSHSHSHSHSHS 235

|||||

seq_name: gb_est42:AW466480

seq_documentation_block:

LOCUS AW466480 523 bp mRNA EST 24-FEB-2000

DEFINITION f160d02.x1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone 2642019 3' similar to TR:Q92277 Q92277 WILLIAMS-BEUREN SYNDROME

DELETION TRANSCRIPT 9 HOMOLOG. ; mRNA sequence.

ACCESSION AW466480.1 GI:7036586

VERSION EST.

KEYWORDS zebrafish.

SOURCE Danio rerio

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Cyprinidae; Rasbora; Danio.

REFERENCE 1 (bases 1 to 523)

AUTHORS Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,

Hallier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,

Martin,J., Pape,D., Steptoe,M., Underwood,K., Reising,B.,

Ritter,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.

Washu Zebrafish EST Project 1999

Unpublished (1999)

On Aug 21, 1998 this sequence version replaced gi:3706503.

Other ESTs: f160d02.y1

Contact: S. L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami

Sequencing by: Washington University Genome Sequencing Center

Trace considered overall poor quality

Seq primer: T7 ET from Amersham

High quality sequence stop: 1.

Location/Qualifiers

1..523

/organism="Danio rerio"

/strain="AB"

/db_xref="taxon:7955"

/clone_lib="2642019"

/sex="mixed (one male and one female, including

unfertilized eggs)"

/dev_stage="adult"

/lab_host="PH10B (phage resistant)"

/note="Vector: pME18S-FL3; Site_1: DraIII (CACTGCTG);

Site_2: DraIII (CACCATGCTG); 1st strand cDNA was primed

with an oligo(dT) primer [ATGTCCTCTTTTCTTTTCTTTT];

double-stranded cDNA was ligated to a DraIII adaptor

[TGTGCTCTACTG], digested and cloned into distinct DraIII

sites of the pME18S-FL3 vector (5' site CACTGCTG, 3' site

CACCATGCTG). XhoI should be used to isolate the cDNA

insert. Size selection was performed to exclude fragments

<1.5kb. Library constructed by Dr. Sumio Sugano

(University of Tokyo Institute of Medical Science) and

kindly donated by Dr. Koichi Kawakami. Custom primers for

sequencing: 5' end primer CTCTGCTCTMAAAGCTGCG and 3' end

primer CGACCTCGACCTGACGACA.

BASE COUNT 96 a 132 c 137 g 158 t

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x AW466480 ..

Align seg 1/1 to: AW466480 from: 1 to: 523

227 PROPROHSHSHSHSHSHSHS 235

|||||

seq_name: gb_gss1:AQ036917

seq_documentation_block:

LOCUS AQ036917 523 bp DNA GSS 11-JUL-1998

DEFINITION CTR-HSP-2331K2.TF CTR-HSP Homo sapiens genomic clone 2331K2,

genomic survey sequence.

ACCESSION AQ036917

VERSION AQ036917.1 GI:3302749

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 523)

AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Baas,S., Linher,K.,

Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,

Simon,M. and Venter,J.C.

Use of a random BAC End Sequence Database for Sequence-Ready Map

Building (1998)

Unpublished (1998)

Other_GSSs: CTR-HSP-2331K2.TR

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mda@amslab.org

Clones are available from Research Genetics (Info@resgen.com). BAC

end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: M13-21

Class: BAC ends.

Location/Qualifiers

1..523

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="2331K2"

/clone_lib="CTR-HSP"

/sex="Male"

/cell_type="Sperm"

/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:

HindIII"

BASE COUNT 129 a 133 c 87 g 174 t

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x AQ036917/rev ..

Align seg 1/1 to reverse of: AQ036917 from: 1 to: 523


```

seq_name: gb_est14:AW585684
seq_documentation_block:
  LOCUS      AW585684      607 bp      mRNA      EST      20-MAR-2000
  DEFINITION  EST317307 MHAM Medicago truncatula/Gloms versiforme mixed EST
  library cDNA clone PMHAM-39K11, mRNA sequence.
  ACCESSION   AW585684
  VERSION     AW585684.1
  KEYWORDS    GI:7265198
  SOURCE      EST.
  ORGANISM    Medicago truncatula/Gloms versiforme mixed EST library.
  AUTHORS     Medicago truncatula/Gloms versiforme mixed EST library
              Eukaryota; mixed EST libraries.
  TITLE       1 (bases 1 to 607)
              Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D.,
              Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
              ESTs from roots of Medicago truncatula after colonization with
              Gloms versiforme
  JOURNAL     Unpublished (2000)
  COMMENT     On May 18, 1998 this sequence version replaced gi:318485.
              Contact: Maria J. Harrison
              Plant Biology Division
              The Samuel Roberts Noble Foundation
              2510 Sam Noble Parkway, Ardmore, OK 73401, USA
              Tel: 580-223-5810
              Fax: 580-221-7380
              Email: mjharrison@noble.org
              Noble EST name:N254612e
              r1cr sequence name:MTDAE66TK
              More information is available at:
              'http://chryslr.tamu.edu/medicago'
              Seq primer: SKmod (CTA gaa cga gga gat cc).
  FEATURES    Location/Qualifiers
              1..607
                /organism="Medicago truncatula/Gloms versiforme mixed EST
                library"
                /cultivar="Medicago truncatula genotype A17"
                /db_xref="taxon:119092"
                /clone="PMHAM-39K11"
                /clone_lib="MHAM"
                /tissue-type="roots colonized with Gloms versiforme"
                /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
                post-inoculation with Gloms versiforme. The library was
                made from a mixture of RNA from each of these stages."
                /lab_host="E. coli strain SOLR"
                /note="Vector: pluescript SK-; Site_1: EcoRI; Site_2:
                XhoI; cDNA was prepared from polyA+ enriched RNA from
                roots harvested at 10, 17, 22, 31 and 38 days
                post-inoculation with Gloms versiforme. The cDNA was
                directionally ligated into the UniZap XR vector from
                Stragene and packaged using Gigapack III Gold packaging
                extracts. Plasmids containing cDNA inserts were excised
                from the recombinant lambdaZap phage using Ex-assist
                helper phage and propagated in SOLR cells."
  BASE COUNT      141 a      166 c      118 g      182 t
  ORIGIN
alignment_scores:
  Quality:      9.00      Length:      9
  Ratio:        1.000      Gaps:      0
  Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
  US-09-332-522B-2 x AW585684      ..
  Align seg 1/1 to: AW585684 from: 1 to: 607
  227 ProProProH1sh1sh1sh1sh1sh1s 235
  ||||||||||||||||||||||||||||
  145 CCGCCACGCCATCACCACACACACAC 171
seq_name: gb_est25:A1814683

```

```

seq_documentation_block:
  LOCUS       AI814683             666 bp             mRNA             EST             07-MAR-2000
  DEFINITION   w66b09.r1 NCI_CGAP_Pan1 Homo sapiens CDNA clone IMAGE:2420345 3'
  similar to contains element MSRI repetitive element ; , mRNA
  sequence.
  ACCESSION   AI814683
  VERSION     AI814683.1      GI:5425898
  KEYWORDS
  SOURCE      human.
  ORGANISM    Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  REFERENCE   1 (bases 1 to 666)
               NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
               National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
               Tumor Gene Index
               Unpublished (1997)
               On Jun 5, 1998 this sequence version replaced gi:3188156.
               Contact: Robert Strausberg, Ph.D.
               Tel: (301) 496-1550
               Email: Robert_Strausberg@nih.gov
               Life Technologies catalog #: 11548-013
               DNA sequencing by: Washington University Genome Sequencing Center
               clone distribution: NCI-CGAP clone distribution information can be
               found through the I.M.A.G.E. Consortium/LNL at:
               www-bio.llnl.gov/bbrp/image/image.html
               Insert Length: 2397 Std Error: 0.00
               Seq primer: -400P from Gldco
               High quality sequence stop: 373.
               Location/Qualifiers
                 1..666
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone_image="2420345"
                   /clone_id="NCI_CGAP_Pan1"
                   /issue_type="adenocarcinoma"
                   /lab_host="DH10B"
                   /note="Organ: pancreas; Vector: pCMV-Sport6; Site:1: SalI
                   Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                   Average insert size 1.72 kb. Life Technologies catalog #:
                   11548-013"
  BASE COUNT      82 a      102 c      272 g      148 t      62 others
  ORIGIN
alignment_scores:
  Quality:      9.00      Length:      9
  Ratio:        1.000      Gaps:      0
  Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
  US-09-332-522B-2 x AI814683/rev ..
Align seg 1/1 to reverse of: AI814683 from: 1 to: 666
227 PROPRPRTSHSHSHSHSHSHS 235
|||||
381 CCCCACCCCATCATCACCACATCAC 355
seq_name: gb_gss12:A0866173
seq_documentation_block:
  LOCUS       A0866173             779 bp             DNA             GSS             03-NOV-1999
  DEFINITION   nbebd0027E12f CUGI Rice BAC Library (EcoRI) Oryza sativa
  genomic clone nbebd0027E12f, genomic survey sequence.
  ACCESSION   A0866173
  VERSION     A0866173.1      GI:5216630
  KEYWORDS
  SOURCE      GSS.
               Oryza sativa.
               Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta.

```

REFERENCE
1 (bases 1 to 779)
Magnoliophyta: Liliopsida: Poales; Poaceae: Oryza.
AUTHORS
Wang, R.A. and Dean, R.A.
TITLE
A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL
Unpublished (1998)
COMMENT
On Dec 15, 1999 this sequence version replaced gi:4575356.

CONTACT: Wang RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGCG
Class: BAC ends
High quality sequence start: 27
High quality sequence stop: 247.
Location/Qualifiers
1..779
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="npbe0027E12f"
/clone_lib="CUGR-Rice BAC Library (EcoRI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: PBACindigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa,
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 Kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9%. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 142 a 225 c 216 g 195 t 1 others
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x A0866173/rev ..

Align seg 1/1 to reverse of: A0866173 from: 1 to: 779

227 PROPPROHSHSHSHSHSHSHS 235
|||||
284 CCACCTCCTCACATCACCACCAT 258

seq_name: gb_est20:AI415613

seq_documentation_block:
LOCUS AI415613 98 bp mRNA EST 09-FEB-1999
DEFINITION mc59603.x1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
clone IMAGE:352828 3', mRNA sequence.
ACCESSION AI415613

VERSION AI415613.1 GI:4259117
EST.
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
1 (bases 1 to 98)
Mammalia: Eutheria: Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Eukaryota: Metazoa: Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS
Marr, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wyle, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuyk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
CONTACT: Maria W/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

COMMENT
This clone is available royalty-free through LNL; contact the
IMAGS Consortium (info@image.llnl.gov) for further information.
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end.

FEATURES
source Location/Qualifiers
1..98
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:352828"
/clone_lib="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAGTGGAGCGCGCGGAAATTTTCTTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
p7T3D vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT 29 a 21 c 29 g 19 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2 x AI415613 ..

Align seg 1/1 to: AI415613 from: 1 to: 98

191-SeqInProTyThrGua1aMet 198
|||||
46 TCACAGCGCTTACACGAGCAATG 69

seq_name: gb_est14:AA947344

seq_documentation_block:
LOCUS AA947344 134 bp mRNA EST 04-MAY-1998
DEFINITION oq49h11.s1 NCI_CGAP_Kids Homo sapiens cDNA clone IMAGE:1589733 3',
mRNA sequence.
ACCESSION AA947344
VERSION AA947344.1 GI:3108597

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE 1 (bases 1 to 134)
NCT-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 17, 1998 this sequence version replaced g1:1901091.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 76.
Location/Qualifiers
1. 134
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1589733"
/clone_lib="NCI CGAP_K1d5"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACGTGAGAAATTCGGCGCGCAATATTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 39 a 14 c 17 g 64 t
ORIGIN

alignment_scores: Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x AA947344/rev ..
Align seg 1/1 to reverse of: AA947344 from: 1 to: 134

963 ThrlslysserleugluSerleu 970
|||||
68 ACAAAAATCCTTAGAATCATTA 45

seq_name: gb_est14:AA992534

seq_documentation_block:
LOCUS AA992534 136 bp mRNA EST 27-DEC-1998
DEFINITION ct68h03.s1 Soares_total.fetus.Nb2HF8_9w Homo sapiens cDNA clone
IMAGE:1621973 3', mRNA sequence.
ACCESSION AA992534
VERSION AA992534.1 GI:3179290
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 136)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced g1:2151536.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert length: 1239 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.
Location/Qualifiers
1. 136
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1621973"
/clone_lib="Soares_total.fetus.Nb2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCGCCTTAATTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 51 a 17 c 19 g 49 t
ORIGIN

alignment_scores: Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2 x AA992534/rev ..
Align seg 1/1 to reverse of: AA992534 from: 1 to: 136

963 ThrlslysserleugluSerleu 970
|||||
44 ACAAAAATCCTTAGAATCATTA 21

seq_name: gb_est24:A1690336

seq_documentation_block:
LOCUS A1690336 136 bp mRNA EST 16-DEC-1999
DEFINITION tz15h12.x1 NCI-CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2286711 3'
similar to gb:D90209 DNA-BINDING PROTEIN TAXREB7 (HUMAN);, mRNA
sequence.
ACCESSION A1690336
VERSION A1690336.1 GI:4901630
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 136)
NCT-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 15, 1998 this sequence version replaced g1:3221937.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.

OM of: US-09-332-522b-2_copy_335_428 to: GenBank: * out-format : pfs

Date: Sep 27, 2000 3:07 PM

About: Results were produced by the GenCone software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-framer-p2n.model -DEV-xlp
-Q/cg92.1/USFTO.spool/US09332522/runatc.27092000_091221.6535/app_query.fasta-1.1355
-DB-GenBank -OFMT-fastap -SUFFIX-011.rge -GAP-4.500
-GAPEXT-0.050 -MINMATCH-0.100 -LOOPCL-0.000 -LOOPEXT-0.000
-OGAP-4.500 -OGAPEXT-0.050 -XGAP-60.000 -XGAPEXT-60.000
-FGAP-6.000 -FGAPEXT-7.000 -YGAP-60.000 -YGAPEXT-60.000
-DELOP-6.000 -DELEXT-7.000 -START-1 -MATRIX-oligo
-TRANS-human40.ccl -LIST-45 -DOCALIGN-200 -THR.SCORE-quality
-THR.MN-1 -ALIGN-45 -MODE-LOCAL -OUTFM-pts -NORM-ext -MINLEN-0
-MAXLEN-2000000000 -USER=US09332522.8CGN1.1.6998 -NCP-6
-ICPU-3 -LONGLOG -NO_XLPXY -WAIT -THREADS-1

Search information block:

Query: US-09-332-522b-2_copy_335_428
Query length: 94
Database: GenBank: *
Database sequences: 972840
Database length: 892348106
Search time (sec): 3472.210000

WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

Score_list:

Sequence	strd Orig	ZScore	EScore	len	Document
gb_in1:CEV47D3B	-	77.00	1417.83	7.1e+71	95968 AL031635 Caenorhabditis elegans
gb_in1:CEV66A7	-	77.00	1413.17	1.3e+70	197735 A0282282 Caenorhabditis elegans
gb_in1:CEV7D3	-	77.00	1409.72	2.0e+70	337565 A028865 Caenorhabditis elegans
gb_in1:CEH10N3	-	42.00	757.65	4.2e+34	36780 Z94157 Caenorhabditis elegans
gb_in1:CEH10N3	+	14.00	224.56	0.0002	84339 Z94157 Caenorhabditis elegans
gb_in1:CEH10N3	+	9.00	124.02	82.26	101215 AC006966 Homo sapiens PAC cl
gb_in1:AC006966	+	9.00	122.84	95.66	101215 AC006966 Homo sapiens PAC cl
gb_in1:AC006966	+	9.00	119.44	148.00	171592 AC009705 Homo sapiens chromo
gb_in1:HS124F	+	8.00	141.57	8.66	289 Z5667 H. sapiens CPG island DN
gb_in1:AC020280	+	8.00	106.82	747.21	63442 AC020280 Drosophila melanog
gb_in1:AC011183	+	8.00	103.99	830.16	72056 AC011183 Homo sapiens clone 1
gb_in1:HS324L9	-	8.00	103.87	73427	84339 A023713 Human DNA sequence f
gb_in1:AC005985	-	8.00	103.03	939.05	83639 AC005985 Drosophila melanog
gb_in1:AC005985	+	8.00	104.87	958.67	85757 AC005985 Drosophila melanog
gb_in1:AL136180	+	8.00	103.92	1.1e+03	99487 AL136180 Homo sapiens chromo
gb_in1:AL137139	+	8.00	103.39	1.2e+03	107860 AL137139 Homo sapiens chromo
gb_in1:HS0399E4	+	8.00	103.22	1.2e+03	110771 AL121948 Homo sapiens chromo
gb_in1:AC009557	+	8.00	102.26	1.3e+03	128550 AC009557 Homo sapiens chromo
gb_in1:AL157902	+	8.00	101.61	1.5e+03	142249 AL157902 Homo sapiens chromo
gb_in1:AL136133	+	8.00	101.60	1.5e+03	142394 AL136133 Homo sapiens chromo
gb_in1:AL136133	+	8.00	101.34	1.5e+03	148298 AC011350 Homo sapiens chromo
gb_in1:AC024178	+	8.00	100.37	1.7e+03	167272 AC024178 Homo sapiens chromo
gb_in1:HS121G13	+	8.00	100.19	1.7e+03	176932 Z86062 Human DNA sequence fr
gb_in1:AC007406	+	8.00	100.19	1.7e+03	177402 AC007406 Homo sapiens clone
gb_in1:AL139019	+	8.00	100.09	1.8e+03	180140 AL139019 Homo sapiens chromo
gb_in1:AC012347	+	8.00	100.08	1.8e+03	180435 AC012347 Homo sapiens clone
gb_in1:AC020581	+	8.00	99.95	1.8e+03	184152 AC020581 Homo sapiens chromo
gb_in1:AC011638	+	8.00	99.94	1.8e+03	184328 AC011638 Homo sapiens clone
gb_in1:AC016721	+	8.00	99.91	1.8e+03	185286 AC016721 Homo sapiens chromo
gb_in1:AC008532	+	8.00	99.80	1.8e+03	188316 AC008532 Homo sapiens chromo
gb_in1:AC021134	+	8.00	99.72	1.8e+03	189184 AC021134 Homo sapiens chromo
gb_in1:AE001572	+	8.00	99.67	1.9e+03	193574 AE001574 Drosophila melanog
gb_in1:AC009329	+	8.00	99.49	1.9e+03	193530 AC009329 Drosophila melanog
gb_in1:AC008591	+	8.00	99.44	1.9e+03	193919 AC008591 Homo sapiens chromo
gb_in1:AC026782	+	8.00	98.98	2.1e+03	213902 AC026782 Homo sapiens chromo
gb_in1:AL139331	+	8.00	98.59	2.1e+03	227106 AL139331 Homo sapiens chromo
gb_in1:AE003674	+	8.00	96.84	2.7e+03	298028 AE003674 Drosophila melanog
gb_in1:AC025442	+	8.00	96.41	2.8e+03	318640 AC025442 Homo sapiens clone
gb_in1:ISA248361	+	7.00	124.84	74.00	202 AJ248361 Homo sapiens STS N1-2

gb_in1:CVU18251	-	7.00	123.27	90.60	258	U18251 Chortophaga viridif
gb_in1:CVU18252	-	7.00	123.27	90.60	258	U18252 Chortophaga viridif
gb_in1:AF145491	-	7.00	123.27	90.60	258	AF145491 Schistosoma gregor
gb_in1:AF028148	-	7.00	122.97	94.07	270	AF028148 Rattus norvegicus
gb_in1:AF028148	-	7.00	122.15	104.61	307	M97282 Sciarus tamiscarius
gb_in1:AF028148	-	7.00	121.28	116.86	351	AJ237293 Lama glama rearing

seq_name: gb_in1:CEV47D3B

seq_documentation_block:

LOCUS CEV47D3B 95968 bp DNA INV 14-DEC-1999
DEFINITION Caenorhabditis elegans cosmid Y47D3B, complete sequence.
ACCESSION AL031635 298865
VERSION AL031635.1 GI:3646936
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditioidea; Rhabditidae; Pelodierinae; Caenorhabditis.
1 (bases 1 to 95968)

REFERENCE

AUTHORS

TITLE

Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)

JOURNAL

MEDLINE

REMARK

REFERENCE

AUTHORS

TITLE

JOURNAL

Direct Submission
Submitted (22-SEP-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RU, England and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jesse@sanger.ac.uk or rwenematode.wustl.edu
Coding sequences below are predicted from computer analysis, using
predictions from Genefinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:-
http://webc.sanger.ac.uk/cgi-
bin/displaydb-wormaccsclass-Sequence-objct-Y47D3B
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone Y47D3B.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone H10N23 is at 61856 in this sequence. The
true right end of clone Y47D3B is at 104 in this sequence. The start
of this sequence (1..104) overlaps with the end of sequence 281134.
The end of this sequence (95863..95968) overlaps with the start of
sequence AL032622.
Location/Qualifiers

FEATURES

source

1..95968
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="Y47D3B"
7188..12736
/gene="Y47D3B.1"

gene

CDS

join(7188..7236,7287..7436,7488..7605,7654..7713,
8884..9277,10137..10318,11510..11678,11731..11860,
12379..12481,12657..12736)
/gene="Y47D3B.1"

[illegible]

PRHOYMSNRKRSKSTSTKCYGKSFSTRSRELFCLIFA"
 34913..48475
 /gene="Y47D3B.5"
 join(34913..35016,38489..38669,39530..39640,39697..39839944..40076,40805..40983,41034..41172,42063..42283,42356..42498,42550..42774,43192..43259,43308..43458,43538..43666,44148..44300,44351..44419,44474..44658,44702..44801,445577..45675,45727..46265,47093..47246,47606..47707,48365..48475)
 /gene="Y47D3B.5"
 /note="CDNA EST YK449e10.5 comes from this gene; CDNA EST EMBL:D5319 comes from this gene; CDNA EST EMBL:C12322 comes from this gene; CDNA EST EMBL:D3263 comes from this gene"
 /codon_start=1
 /protein_id="CAA21041.1"
 /db_xref="GI:3881007"
 /db_xref="SPTREMBL:O9X01"
 /translation="MPNPSEPTPIQASFTTLLDGYTDALPEKYNDFNFRSPYSRSLNSTSLGSKVYRFQSFEQSRDLTTIDDEEELRYKPAPEPTFKLQVSRITVRMLA VGFSEKAGSLGVEGTGIVGPMPSIFSLGCKSIYKXKADRIASATLVLVYFLGMW VFLFEDHYREVKLKEKRLDRLMRLNRTFPIRSMPYRQOGLNASEVLEIDYKELKLV VRLPECLMDMYGALFVGLTFLTTIGTIGVPTAIGRAASVYAIYVGLVLAITLISK CGKMTDLSSEKQOQRHIOITTEKAKTKNRLRKTKIKSELYEANTGACADEKREK VESSTPIWLALICVYVAVGCSLFLMELNTRTFSLFELTISLTIGLGVIPK PHNEIVFVLIIVGLSIVSMFISVVOQIIEWLCRMIEQIKREYMKMGALVDEEIKR IFANDPLSFVAPARIMGSEQNHLKDIKDFDLLESEKIQEIPYKAGATQVEYEQ SMACDPMSNHLKPAEHOGTQMSRVDETGGPKNVELEKLRQDPAITSRSNPDSI SDATSLPDSMSVPAKRRKRNTEKVCOCDDPMGTAFFKRDATGATQDLOFOIDE IAKIANLQTORVRPIYEMATSNAPMDSPLNSEALDIPGLSQADINVICEMSSSF FKSQELAQIADIGVKDYETIFSDSQSLTSLPLVPKRNIAITSVSLVLEDLKA SKSKSRKSISSPILPENSJLHEETQIPLQNNVHMATDCTSLVHRSVTSPVGEK DAETEMPOKMDRWVSFVSTAIKDONTSPVEQTPDSDRIQOTSISEMGLKKEED ESQOTSLCIDVYGSKKEGKKKASRENQASLSGADVTSSNPSTISDSEKLNHK IKTLLAPNLTASVMSSTOKSPPGFGRFSEISTARSAVYIATEDRQV IYOTDSDYLKTIARLDEYRNKTKQFLPVCAALALESEVYEPFSDGRSSEKSHYITFGG RRSLSGGRKKPRQDMRTGESQTQSADALLEVSSEKSRISIPRPNCKKGVLSR HPSLPTGISRGKVDYVIOHEKGIHNPAIROSPIKTIROYSLDHVI"
 51392..52342
 /gene="Y47D3B.6"
 join(51392..51449,51544..51641,51689..52342)
 /gene="Y47D3B.6"
 /note="predicted using GeneFinder: CDNA EST YK506e5.3 comes from this gene"
 /codon_start=1
 /protein_id="CAA21043.2"
 /db_xref="GI:5824793"
 /db_xref="GI:5824793"
 /translacion="MRLITTCATLLAFTAAQVYVADYLYSCASAVPCAPACQASVC ITPAAAPMMAMPAPAPAPAPAPLPAPAPAPAPVFAAPAPAPAPAPAPAPV IAQAPGCAAPCIGAOCAGACACAPNVAIVSVYICGPAIPAPVAPVARRLLIPO ALGSPCEPVECTGGSCVGCCLCPSELIOGTGVCATVATIGVPPVPIAPVAPV VAIAGPAPAPVAPVSAAPACVAGVAGVCSAGVCGAGVYPLDQACVRRLL"
 /complement(52583..56730)
 /gene="Y47D3B.7"
 complement(join(52583..52774,52830..53190,61122..61275,61322..61549,61609..62005,62315..62843,62889..63082,63606..63841,64186..64605,64659..64767,65539..65848,66107..66239,66652..66730))
 /gene="Y47D3B.7"
 /note="predicted using GeneFinder: similar to Helix-loop-helix DNA-binding domain; CDNA EST EMBL:D34815 comes from this gene; CDNA EST EMBL:D35004 comes from this gene; CDNA EST EMBL:D68853 comes from this gene; CDNA EST EMBL:D67853 comes from this gene; CDNA EST YK49h3.5 comes

Align seg 1/1 to reverse of: CEY47D3B from: 1 to: 9568

```

18 Prollythrgluargatgthralahisasnleuileglulyslytyr 34
|||||
63839 CCAAAACCAGGCGTGTACGACACAAATCTCATCGAAAAGATATAG 63790
34 gcysserileasnspargilegninglneulysvalleuencysglya 51
|||||
63789 ATGCTCAATTAATGATGCAATTCACAGCTGAAAGTACTTTGTGTGGGG 63740
51 spglualalysleuserlysseralathrleuargalalegluhis 67
|||||
63739 ATGAGCTAGAGCTTTCAAAATCGGCAACACTAGACGGCTATTGAACAT 63690
68 lileglulvalgluhisgluasnglnvalleulysishisvalglucl 84
|||||
63689 ATCGAGAGAGGTTGACACGACGAAATCAGGTGTGACATCATGTGACACA 63640
84 nmetarglysthrleuglnasnsnargleu 94
|||||
63639 AATGAGAAAGACACTGCAAGATATATCGATT 63609

```

seq_name: gb_htg1:CEY66A7

```

seq_documentation_block:
LOCUS CEY66A7.197735 bp DNA HTG 03-DEC-1998
DEFINITION Caenorhabditis elegans chromosome III clone Y66A7, *** SEQUENCING
IN PROGRESS ***, in unordered pieces.
ACCESSION AL022282
VERSION AL022282.1 GI:3451532
KEYWORDS HTG: HTGS, PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;
Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 197735)
AUTHORS Matthews, L.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rlvnematode.wustl.edu

```

COMMENT

On Aug 25, 1998 this sequence version replaced g1:3250755.
 IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence. Work on the sequence is in progress
 and the release of this data is based on the understanding that the
 sequence may change as work continues. The sequence may be
 contaminated with foreign sequence from E.coli, yeast, vector,
 phage etc. Order of segments is not known; 800 n's separate
 segments.
 * NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES

```

source
1..197735
  /location=Qualifiers
  /organism="Caenorhabditis elegans"
  /db_xref="taxon:6239"
  /chromosome="III"
  /clone="Y66A7"
BASE COUNT 62861 a 35824 c 35278 g 62783 t 989 others
ORIGIN

```

alignment_scores:

Quality: 77.00 Length: 77
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x CEY66A7/rev ..

Align seg 1/1 to reverse of: CEY66A7 from: 1 to: 197735

```

18 Prollythrgluargatgthralahisasnleuileglulyslytyr 34
|||||
4793 CCAAAACCAGGCGTGTACGACACAAATCTCATCGAAAAGATATAG 4744
34 gcysserileasnspargilegninglneulysvalleuencysglya 51
|||||
4743 ATGCTCAATTAATGATGCAATTCACAGCTGAAAGTACTTTGTGTGGGG 4694
51 spglualalysleuserlysseralathrleuargalalegluhis 67
|||||
4693 ATGAGCTAGAGCTTTCAAAATCGGCAACACTAGACGGCTATTGAACAT 4644
68 lileglulvalgluhisgluasnglnvalleulysishisvalglucl 84
|||||
4643 ATCGAGAGAGGTTGACACGACGAAATCAGGTGTGACATCATGTGACACA 4594
84 nmetarglysthrleuglnasnsnargleu 94
|||||
4593 AATGAGAAAGACACTGCAAGATATATCGATT 4563

```

seq_name: gb_htg1:CEY47D3

```

seq_documentation_block:
LOCUS CEY47D3.337565 bp DNA HTG 02-SEP-1999
DEFINITION Caenorhabditis elegans chromosome III clone Y47D3, *** SEQUENCING
IN PROGRESS ***, in unordered pieces.
ACCESSION Z98865
VERSION Z98865.1 GI:4164282
KEYWORDS HTG: HTGS, PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;
Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 337565)
AUTHORS Matthews, L.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1999) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rlvnematode.wustl.edu

```

COMMENT

Order of segments is not known; 800 n's separate segments.
 IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence. Work on the sequence is in progress
 and the release of this data is based on the understanding that the
 sequence may change as work continues. The sequence may be
 contaminated with foreign sequence from E.coli, yeast, vector,
 phage etc.
 * NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES

```

source
1..337565
  /location=Qualifiers
  /organism="Caenorhabditis elegans"
  /db_xref="taxon:6239"
  /chromosome="III"
  /clone="Y47D3"
BASE COUNT 108087 a 59856 c 60930 g 107892 t 800 others
ORIGIN

```

alignment_scores:

Quality: 77.00 Length: 77
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522B-2_COPY_335_428 x CEY47D3/rev ..

Align seg 1/1 to reverse of: CEY47D3 from: 1 to: 337565

```

18  ProlystThrgluArgArGThralaHsAsnLeuIleGluLysLysTyrAr 34
    |||||||
305436  CCAAAAACGAGCGTGTACGGCACATCTCATCGAAAAGAGTATAG 305387
    |||||||
34  gCySerIleAsnAspArgIleGlnInLeuLysValLeuLeuCyGlyA 51
    |||||||
305386  ATGCTCAATAAATGATCGAATTCACACAGCTGAAGTACTTTGTGTGGGG 305337
    |||||||
51  spGluAlaLysSerLysSerAlaThLeuArgArGAlaIleGluHs 67
    |||||||
305336  ATGAAGCTAAGCTTCAAAATCGGCAACACTACGAGCGGCTATTGAACAT 305287
    |||||||
68  IleGluGluValGluHsGluAsnGlnValLeuLysHsIleValGluI 84
    |||||||
305286  ATCGAGAGAGTGTGAACACGACGATCGTGTGAACATCATCTTACACA 305237
    |||||||
84  nMetArgLysThrLeuGlnAsnAsnArgLeu 94
    |||||||
305236  AATGAGAAAGACACTGCAGATAATCATGATTA 305206
    |||||||

seq_name: gb_htg1:CEH10N23

seq_documentation_block:
LOCUS      CEH10N23      36780 bp      DNA      22-SEP-1998
DEFINITION Caenorhabditis elegans chromosome III clone H10N23, *** SEQUENCING
IN PROGRESS ***, in unordered pieces.
ACCESSION  Z94157.1 GI:3377969
VERSION    Z94157.1 GI:3377969
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     Caenorhabditis elegans.
ORGANISM   Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiida;
            Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
            1 (bases 1 to 36780)
REFERENCE  1 (bases 1 to 36780)
AUTHORS    Sulston, J.
TITLE      Direct Submission
JOURNAL    Submitted (21-SEP-1998) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jess@anger.ac.uk or rwenematode.wustl.edu
COMMENT    On Aug 3, 1998 this sequence version replaced g1:1945146.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.

FEATURES
    source             1..36780
                        /organism="Caenorhabditis elegans"
                        /db_xref="taxon:6239"
                        /chromosome="III"
                        /clone="H10N23"

BASE COUNT      8697 a      5380 c      5339 g      8564 t      8800 others
ORIGIN
..
alignment_scores:
    Quality:      42.00      Length:      42
    Ratio:        1.000      Gaps:      0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x CEH10N23/rev ..
Align seg 1/1 to reverse of: CEH10N23 from: 1 to: 36780
53 AlalysLysSerLysSerAlaThrLeuArgArGAlaIleGluHsIleG1 69

```

```

1353  GCTAAGCTTTAAATCGGCAACACTACGACGGCTTTGACATTAATCGA 1304
    |||||||
69  uGluValGluHsGluAsnGlnValLeuLysHsIleValGluInMetA 86
    |||||||
1303  GGAGGTGTACACGAGAAATCAGGTGTGAAGCATCATGTTGAACAAATGA 1254
    |||||||
86  rGlyThrLeuGlnAsnAsnArgLeu 94
    |||||||
1253  GAAAGACACTGCAGATAATCATGATTA 1228
    |||||||

seq_name: gb_htg1:CEH10N23

seq_documentation_block:
LOCUS      CEH10N23      36780 bp      DNA      22-SEP-1998
DEFINITION Caenorhabditis elegans chromosome III clone H10N23, *** SEQUENCING
IN PROGRESS ***, in unordered pieces.
ACCESSION  Z94157.1 GI:3377969
VERSION    Z94157.1 GI:3377969
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     Caenorhabditis elegans.
ORGANISM   Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiida; Rhabditidae;
            Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
            1 (bases 1 to 36780)
REFERENCE  1 (bases 1 to 36780)
AUTHORS    Sulston, J.
TITLE      Direct Submission
JOURNAL    Submitted (21-SEP-1998) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jess@anger.ac.uk or rwenematode.wustl.edu
COMMENT    On Aug 3, 1998 this sequence version replaced g1:1945146.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.

FEATURES
    source             1..36780
                        /organism="Caenorhabditis elegans"
                        /db_xref="taxon:6239"
                        /chromosome="III"
                        /clone="H10N23"

BASE COUNT      8697 a      5380 c      5339 g      8564 t      8800 others
ORIGIN
..
alignment_scores:
    Quality:      14.00      Length:      14
    Ratio:        1.000      Gaps:      0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x CEH10N23 ..
Align seg 1/1 to: CEH10N23 from: 1 to: 36780
18  ProlystThrgluArgArGThralaHsAsnLeuIleGluLys 31
    |||||||
6856  CCAAAAACGAGCGTGTACGGCACATCTCATCGAAAAG 6897
    |||||||

seq_name: gb_htg9:AC023160

seq_documentation_block:
LOCUS      AC023160      84339 bp      DNA      19-FEB-2000
DEFINITION Homo sapiens chromosome 12p clone RP11-103508, *** SEQUENCING IN
PROGRESS ***, 66 unordered pieces.

```


ACCESSION	AC023160
VERSION	AC023160.3
KEYWORDS	GTG; HTGS_PHA5EL
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 84339)
AUTHORS	Muzdy,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K., Bouda,B., Bouck,J., Boyle,S., Brooks,A., Bulay,C., Bunac,C., Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C., David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Frantz,P., Ganeish,R., Gorrell,J.H., Gorrell,L.L., Guenard,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M., Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondejowski,N., Kong,Y., Kovar,C., Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mel,G., Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,S., Nguyen,S., Ouellet,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L., Qullis,M., Reller,D., Rives,M., Samuel,S., Say,D., Scherer,S., Sha,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugrue,R., Taber,P., Taylor,T., Vasquez,L., Vanson,R., Vo,O., Wabhang,R., Wellington,S., Weinstein,G., Weinstein,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 84339)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (09-FEB-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Feb 19, 2000 this sequence version replaced gi:6978335.
	----- Genome Center -----
	Center: Baylor College of Medicine
	Center code: BCM
	Web site: http://www.hgsc.bcm.tmc.edu/
	Contact: hgsc-help@bcm.tmc.edu
	----- Project Information -----
	Center project name: HAEU
	Center clone name: RP11-1035D8
	----- Summary Statistics -----
	Sequencing vector: M13; L08821
	Chemistry: Dye-terminator Big Dye; 99% of reads
	Assembly program: Phrap; version 0.940611
	Consensus quality: 39046 bases at least Q40
	Consensus quality: 53728 bases at least Q30
	Consensus quality: 61584 bases at least Q20
	Estimated insert size: 60193; sum-of-coverage estimation
	Quality coverage: 0.9x in Q20 bases; sum-of-coverage estimation

	* NOTE: This is a 'working draft' sequence. It currently
	* consists of 66 contigs. The true order of the pieces
	* is not known and their order in this sequence record is
	* arbitrary. Gaps between the contigs are represented as
	* runs of N, but the exact sizes of the gaps are unknown.
	* This record will be updated with the finished sequence
	* as soon as it is available and the accession number will
	* be preserved.

	* 1 781: contig of 781 bp in length
	* 782 801: gap of unknown length
	* 802 1577: contig of 776 bp in length
	* 1578 1587: gap of unknown length
	* 1598 2250: contig of 653 bp in length
	* 2251 2270: gap of unknown length
	* 2271 3037: contig of 767 bp in length
	* 3038 3057: gap of unknown length
	* 3058 3805: contig of 748 bp in length
	* 3806 3825: gap of unknown length
	* 3826 4589: contig of 764 bp in length

4590	4609	gap of unknown length
4610	5375	contig of 766 bp in length
5376	5395	gap of unknown length
5396	6178	contig of 783 bp in length
6179	6598	gap of unknown length
6199	6668	contig of 770 bp in length
6669	6688	gap of unknown length
6989	7739	contig of 751 bp in length
7740	7759	gap of unknown length
7760	8497	contig of 738 bp in length
8498	8517	gap of unknown length
8518	9295	contig of 778 bp in length
9296	9315	gap of unknown length
9316	10064	contig of 749 bp in length
10065	10084	gap of unknown length
10085	10826	contig of 742 bp in length
10827	10846	gap of unknown length
10847	11593	contig of 747 bp in length
11594	11613	gap of unknown length
11614	12459	contig of 846 bp in length
12460	12479	gap of unknown length
12480	13414	contig of 935 bp in length
13415	13434	gap of unknown length
13435	14480	contig of 746 bp in length
14181	14201	gap of unknown length
14201	15368	contig of 1168 bp in length
15369	15388	gap of unknown length
15389	16736	contig of 1348 bp in length
16737	16756	gap of unknown length
16757	18201	contig of 1445 bp in length
18202	18221	gap of unknown length
18222	19403	contig of 1182 bp in length
19403	19423	gap of unknown length
19424	20283	contig of 860 bp in length
20284	20303	gap of unknown length
20304	21471	contig of 1168 bp in length
21472	21491	gap of unknown length
21492	22510	contig of 1019 bp in length
22511	22530	gap of unknown length
22531	23589	contig of 1059 bp in length
23590	23609	gap of unknown length
23610	24381	contig of 772 bp in length
24382	24401	gap of unknown length
24402	25283	contig of 882 bp in length
25284	25304	gap of unknown length
25304	26220	contig of 917 bp in length
26221	26240	gap of unknown length
26241	27311	contig of 971 bp in length
27312	27331	gap of unknown length
27332	28567	contig of 1336 bp in length
28568	28587	gap of unknown length
28588	29637	contig of 1050 bp in length
29638	29657	gap of unknown length
29658	30624	contig of 967 bp in length
30625	30644	gap of unknown length
30645	31412	contig of 768 bp in length
31413	31432	gap of unknown length
31433	32535	contig of 1203 bp in length
32636	32655	gap of unknown length
32656	33615	contig of 960 bp in length
33616	33635	gap of unknown length
33636	34551	contig of 916 bp in length
34552	34571	gap of unknown length
34572	35771	contig of 1200 bp in length
35772	35791	gap of unknown length
35792	37668	contig of 1877 bp in length
37669	37688	gap of unknown length
37689	38794	contig of 1106 bp in length
38795	38814	gap of unknown length
38815	39769	contig of 955 bp in length
39770	39789	gap of unknown length
39790	41365	contig of 1576 bp in length
41366	41385	gap of unknown length

```

* 41386 43097: contig of 1712 bp in length
* 43098 43117: gap of unknown length
* 44505 44505: contig of 1388 bp in length
* 44506 44525: gap of unknown length
* 44526 46343: contig of 1818 bp in length
* 46344 46363: gap of unknown length
* 46364 47498: contig of 1135 bp in length
* 47499 47519: gap of unknown length
* 47519 49233: contig of 1715 bp in length
* 49234 49253: gap of unknown length
* 49254 50558: contig of 1305 bp in length
* 50559 50578: gap of unknown length
* 52316 52316: contig of 1738 bp in length
* 52317 52336: gap of unknown length
* 52337 54073: contig of 1737 bp in length
* 54074 54093: gap of unknown length
* 54094 55230: contig of 1137 bp in length
* 55231 55250: gap of unknown length
* 55251 57008: contig of 1758 bp in length
* 57009 57029: gap of unknown length
* 57029 58615: contig of 1587 bp in length
* 58616 58636: gap of unknown length
* 58636 60102: contig of 1467 bp in length
* 60103 60122: gap of unknown length
* 60123 61513: contig of 1391 bp in length
* 61514 61533: gap of unknown length
* 61534 62895: contig of 1361 bp in length
* 62895 62914: gap of unknown length
* 62915 64661: contig of 1747 bp in length
* 64662 64681: gap of unknown length
* 64682 66480: contig of 1799 bp in length
* 66481 66500: gap of unknown length
* 66501 67770: contig of 1270 bp in length
* 67771 67791: gap of unknown length
* 67791 70867: contig of 3077 bp in length
* 70868 70887: gap of unknown length
* 70888 72480: contig of 1593 bp in length
* 72481 72501: gap of unknown length
* 72501 74067: contig of 1566 bp in length
* 74067 74087: gap of unknown length
* 74087 76594: contig of 2508 bp in length
* 76595 76614: gap of unknown length
* 76615 78851: contig of 2237 bp in length
* 78852 78871: gap of unknown length
* 78872 81371: contig of 2500 bp in length
* 81372 81391: gap of unknown length
* 81392 84339: contig of 2948 bp in length.
FEATURES
      source
      1..84339
      /organism="Homo sapiens"
      /db_xref="taxon:9606"

```

```

alignment_scores:
  quality: 9.00      length: 9
  ratio: 1.000      gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

```

```

alignment_block:
  us-09-332-522b-2_copy_335_428 x AC023160/rev ..

```

```

Align seg 1/1 to reverse of: AC023160 from: 1 to: 84339

```

```

54 LysleuSerlysserAlaThrLeuArg 62
|||||
27384 AAGCTGTCAAGAGTGCACCTTAGG 27358

```

```

seq_name: gb_pt4:AC006966

```

```

seq_documentation_block:
LOCUS      AC006966      101215 bp      DNA      21-DEC-1999
DEFINITION Hom sapiens PAC clone RP4-563H24 from 7q34-q36, complete sequence.
ACCESSION  AC006966
VERSION    AC006966.3  GI:5091657

```

```

KEYWORDS
SOURCE      HTG.
ORGANISM    human.
REFERENCE   Homo sapiens
AUTHORS     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       1 (bases 1 to 101215)
JOURNAL     Sulston, J.E. and Waterston, R.
            Toward a complete human genome sequence
            Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE     99063792
REFERENCE   2 (bases 1 to 101215)
AUTHORS     Walker, C., Scott, K., Bauer, C. and Harkins, R.
TITLE       The sequence of Homo sapiens PAC clone RP4-563H24
JOURNAL     Unpublished
REFERENCE   3 (bases 1 to 101215)
AUTHORS     Waterston, R.H.
TITLE       Direct Submission
JOURNAL     Submitted (05-MAR-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4 (bases 1 to 101215)
AUTHORS     Waterston, R.H.
TITLE       Direct Submission
JOURNAL     Submitted (17-JUN-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5 (bases 1 to 101215)
AUTHORS     Waterston, R.H.
TITLE       Direct Submission
JOURNAL     Submitted (22-OCT-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Jun 17, 1999 this sequence version replaced gi:4508153.
REFERENCE   6 (bases 1 to 101215)
AUTHORS     Waterston, R.
TITLE       Direct Submission
JOURNAL     Submitted (21-DEC-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Jun 17, 1999 this sequence version replaced gi:4508153.
COMMENT

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NIGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@ngri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
This clone was derived from human PAC library RPI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://dacpac.med.buffalo.edu>) using the method described by

Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-1127D14; the clone sequenced to the right is RP4-555L14, 200 bp overlap. Actual start of this clone is unknown; actual end is at base position 20137 of RP4-555L14.

FEATURES

```

Source          1..101215
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="7"
                /map="7q34-q36"
                /clone="RP4-553H24"
                /clone_1lb="RPCI-5"
                34..303
                /rpt_family="Alu"
                304..351
                /rpt_family="L1"
                356..367
                /rpt_family="Alu"
                368..393
                /rpt_family="(CAAAA)n"
                394..506
                /rpt_family="Alu"
                592..864
                /rpt_family="L1"
                891..964
                /rpt_family="Alu"
                945..992
                /rpt_family="Alu"
                995..1128
                /rpt_family="Alu"
                1129..1439
                /rpt_family="Alu"
                1440..1567
                /rpt_family="Alu"
                1618..1761
                /rpt_family="L1"
                1796..1857
                /rpt_family="MER2_type"
                1858..2172
                /rpt_family="Alu"
                2173..2483
                /rpt_family="MER2_type"
                2481..2564
                /rpt_family="MER2_type"
                3122..3428
                /rpt_family="Alu"
                3429..3454
                /rpt_family="(TAAA)n"
                3455..3737
                /rpt_family="Alu"
                3739..3868
                /rpt_family="Alu"
                4373..4551
                /rpt_family="L1"
                4588..4657
                /rpt_family="(GGGA)n"
                4658..4930
                /rpt_family="Alu"
                4961..5205
                /rpt_family="L2"
                5308..5503
                /rpt_family="L1"
                5573..5637
                /rpt_family="(CA)n"
                6052..6487
                /rpt_family="L1"

```

```

repeat_region 6488..6776
               /rpt_family="L1"
misc_feature 7397..7608
             /note="match to EST R79638 (NID:9855919) y189D07.r1"
repeat_region 7860..8102
               /rpt_family="Alu"
               8117..8418
               /rpt_family="Alu"
               8419..8629
               /rpt_family="Alu"
               8788..9120
               /rpt_family="MER2_type"
               9866..10136
               /rpt_family="Retroviral"
               10854..11179
               /rpt_family="Alu"
               11209..11511
               /rpt_family="Alu"
               11524..11813
               /rpt_family="Alu"
               12590..12624
               /rpt_family="AT-rich"
               12864..13026
               /rpt_family="Alu"
               13108..13395
               /note="match to EST F08024 (NID:9677527)"
               13109..13579
               /note="match to EST H12942 (NID:9877762) y170905.r1"
               13338..13592
               /rpt_family="L2"
               13595..14054
               /rpt_family="MALR"
               14067..14458
               /note="match to EST R56473 (NID:9826579) y933f09.s1"
               14114..14451
               /note="match to EST H05531 (NID:9869083) y170905.s1"
               14344..14623
               /rpt_family="Alu"
               15064..15370
               /rpt_family="Alu"
               15724..15842
               /rpt_family="Alu"
               15843..15948
               /rpt_family="Alu"
               15949..16250
               /rpt_family="Alu"
               16251..16455
               /rpt_family="Alu"
               16466..16531
               /rpt_family="AT-rich"

```

```

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

alignment_block:

US-09-332-522b-2_COPY-335_428 x AC006966 ..

Align seg 1/1 to: AC006966 from: 1 to: 101215

5 GluThrMetSerGlnGlyThrThrIle 13

42821 GAAACATGAGCCAGTACCACTATA 42847

seq_name: gb_hg28:AC009705

seq_documentation_block:

LOCUS AC009705 171592 bp DNA HTG 20-APR-2000
 DEFINITION Homo sapiens chromosome 8 clone RP11-248D6 map 8, WORKING DRAFT
 ACCESSION AC009705
 VERSION AC009705.3 GI:7622351

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.

SOURCE human

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 171592)

ATTNORS Bliren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 8, clone RP11-248D6

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 171592)

Bliren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckert,R., Benn,J., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArrellano,K., Depayre,E., Devon,K., Dewar,K., Donegan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Hagos,B., Heathcote,A., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Margulis,N., McKean,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J., Moll,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

Direct Submission

Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 20, 2000 this sequence version replaced gi:5932594.

All repeats were identified using RepeatMasker:

Slit, A.F.A. & Green, P. (1996-1997) RepeatMasker.html

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: L2364

Center clone name: 248_D_6

----- Summary Statistics

Sequencing vector: M13: M77815: 100% of reads

Chemistry: Dye-terminator Big Dye, 100% of reads

Assembly program: Phrap, version 0.960731

Consensus quality: 134976 bases at least Q40

Consensus quality: 151188 bases at least Q30

Consensus quality: 163420 bases at least Q20

Insert size: 170000; agarose-ef

Insert size: 170192; sum-of-ctrls

Quality coverage: 4.3 in Q20 bases; agarose-ef

Quality coverage: 4.3 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently

* consists of 15 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 1030: contig of 1030 bp in length

* 1031 1130: gap of 100 bp

* 1131 2153: contig of 1023 bp in length

* 2154 2253: gap of 100 bp

* 2254 3377: contig of 1124 bp in length

* 3378 3477: gap of 100 bp

* 3478 5255: contig of 1778 bp in length

* 5256 5355: gap of 100 bp

* 5356 7808: contig of 2453 bp in length

* 7809 7908: gap of 100 bp

* 7909 11406: contig of 3498 bp in length

* 11407 11506: gap of 100 bp

* 11507 16884: contig of 5378 bp in length

FEATURES

source

1..171592

16885 16984: gap of 100 bp

* 16985 23335: contig of 6351 bp in length

* 23336 23435: gap of 100 bp

* 23436 28674: contig of 5239 bp in length

* 28675 28774: gap of 100 bp

* 28775 36963: contig of 8189 bp in length

* 36964 37063: gap of 100 bp

* 37064 50783: contig of 13720 bp in length

* 50784 50883: gap of 100 bp

* 50884 67791: contig of 16908 bp in length

* 67792 67891: gap of 100 bp

* 67892 96382: contig of 28491 bp in length

* 96383 96482: gap of 100 bp

* 96483 131712: contig of 35230 bp in length

* 131713 131812: gap of 100 bp

* 131813 171592: contig of 39780 bp in length.

Location/Qualifiers

1..171592

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="8"

/map="8"

/clone="RP11-248D6"

/clone.lib="RP11-11 Human Male BAC"

1..1030

/note="assembly-fragment"

1131..2153

/note="assembly-fragment"

2254..3377

/note="assembly-fragment"

3478..5255

/note="assembly-fragment"

5356..7808

/note="assembly-fragment"

7909..11406

/note="assembly-fragment"

11507..16884

/note="assembly-fragment"

16985..23335

/note="assembly-fragment"

23436..28674

/note="assembly-fragment"

28775..36963

/note="assembly-fragment"

37064..50783

/note="assembly-fragment"

50884..67791

/note="assembly-fragment"

67892..96382

/note="assembly-fragment"

96483..131712

/note="assembly-fragment"

131813..171592

/note="assembly-fragment"

clone_end:17

vector_side:left

BASE COUNT 55500 a 33109 c 32232 g 49322 t 1429 others

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:

US-09-332-522B-2_COPY_335_428 x AC009705 ..

Align seg 1/1 to: AC009705 from: 1 to: 171592

85 Metaglystrleuglinasnasary 93

|||||

105334 ATGAGAAACCCCTTCAAAACATCGA 105360

seq_name: gb_pr1:HS12E4F

seq_documentation_block: 289 bp DNA PRI 18-OCT-1995
LOCUS HS12E4F
DEFINITION H.sapiens Cpg island DNA genomic MseI fragment, clone 12e4, forward
read cpg12e4.ftlc.
ACCESSION 256607
VERSION 256607.1 GI:1027838
KEYWORDS Cpg island; genomic MseI fragment.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 289)
AUTHORS MacDonald,M., Huckie,E., Wilkinson,P. and Mickle,G.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk
2 (bases 1 to 289)
AUTHORS Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
TITLE Purification of Cpg islands using a methylated DNA binding column
JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE 94282070
COMMENT Vector: pGEM-52f(-)
Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
http://www.hgmp.mrc.ac.uk/ for details
or contact: biochip@hgmp.mrc.ac.uk.

FEATURES
source 1..289
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="male"
/dev_stage="adult"
/tissue_type="blood"
/clone_lib="CGI-1"
/clone="12e4"

BASE COUNT 55 a 96 c 86 g 52 t

ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x HS12E4F ..
Align seg 1/1 to: HS12E4F from: 1 to: 289

seq_name: gb_htg6:AC020280

seq_documentation_block: 63442 bp DNA HTG 03-JAN-2000
LOCUS AC020280
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC020280
VERSION AC020280.1 GI:6664617
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 63442)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission

JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10212650 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source 1..63442
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 18372 a 13865 c 13513 g 17692 t

ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x AC020280/rev ..
Align seg 1/1 to reverse of: AC020280 from: 1 to: 63442

seq_name: gb_htg4:AC011183

seq_documentation_block: 72056 bp DNA HTG 01-OCT-1999
LOCUS AC011183
DEFINITION Homo sapiens clone 12_B_11, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC011183
VERSION AC011183.1 GI:6006118
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 72056)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone 12_B_11
JOURNAL Unpublished
2 (bases 1 to 72056)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckert,J.R., Boguslavsky,L., Bouknight,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., DeArrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hatford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karasik,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McMan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tittell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1	898: contig of 898 bp in length
899	gap of unknown length
1785	contig of 887 bp in length
1786	gap of unknown length
2678	contig of 893 bp in length
2679	gap of unknown length
3565	contig of 888 bp in length
3567	gap of unknown length
4474	contig of 908 bp in length
4475	gap of unknown length
5394	contig of 920 bp in length
5395	gap of unknown length
6335	contig of 941 bp in length
6336	gap of unknown length
7293	contig of 958 bp in length
7294	gap of unknown length
8172	contig of 879 bp in length
8173	gap of unknown length
9058	contig of 886 bp in length
9059	gap of unknown length
9962	contig of 904 bp in length
9963	gap of unknown length
10708	contig of 746 bp in length
10709	gap of unknown length
11617	contig of 909 bp in length
11618	gap of unknown length
12521	contig of 904 bp in length
12522	gap of unknown length
13412	contig of 891 bp in length
13413	gap of unknown length
14301	contig of 889 bp in length
14302	gap of unknown length
15220	contig of 919 bp in length
15221	gap of unknown length
16096	contig of 876 bp in length
16097	gap of unknown length
17038	contig of 942 bp in length
17039	gap of unknown length
17964	contig of 926 bp in length
17965	gap of unknown length
18911	contig of 947 bp in length
18912	gap of unknown length
19842	contig of 931 bp in length
19843	gap of unknown length
20736	contig of 894 bp in length
20737	gap of unknown length
21598	contig of 862 bp in length
21599	gap of unknown length
22508	contig of 910 bp in length
22509	gap of unknown length
23395	contig of 887 bp in length
23396	gap of unknown length
24301	contig of 906 bp in length
24302	gap of unknown length
25201	contig of 900 bp in length
25202	gap of unknown length
26120	contig of 919 bp in length
26121	gap of unknown length
27005	contig of 885 bp in length
27006	gap of unknown length
27865	contig of 860 bp in length
27866	gap of unknown length
28756	contig of 891 bp in length
28757	gap of unknown length
29667	contig of 911 bp in length
29668	gap of unknown length
30545	contig of 878 bp in length
	gap of unknown length

30546	31489: contig of 944 bp in length
	gap of unknown length
31490	32445: contig of 956 bp in length
	gap of unknown length
32446	33362: contig of 917 bp in length
	gap of unknown length
33363	34288: contig of 926 bp in length
	gap of unknown length
34289	35195: contig of 907 bp in length
	gap of unknown length
35196	36090: contig of 895 bp in length
	gap of unknown length
36091	36976: contig of 886 bp in length
	gap of unknown length
36977	37702: contig of 726 bp in length
	gap of unknown length
37703	38398: contig of 696 bp in length
	gap of unknown length
38399	39314: contig of 916 bp in length
	gap of unknown length
39315	40226: contig of 912 bp in length
	gap of unknown length
40227	41133: contig of 907 bp in length
	gap of unknown length
41134	42015: contig of 882 bp in length
	gap of unknown length
42016	42879: contig of 864 bp in length
	gap of unknown length
42880	43787: contig of 908 bp in length
	gap of unknown length
43788	44732: contig of 945 bp in length
	gap of unknown length
44733	45645: contig of 913 bp in length
	gap of unknown length
45646	46606: contig of 961 bp in length
	gap of unknown length
46607	47527: contig of 921 bp in length
	gap of unknown length
47528	48457: contig of 930 bp in length
	gap of unknown length
48458	49401: contig of 944 bp in length
	gap of unknown length
49402	50300: contig of 899 bp in length
	gap of unknown length
50301	51179: contig of 879 bp in length
	gap of unknown length
51180	52130: contig of 951 bp in length
	gap of unknown length
52131	53074: contig of 944 bp in length
	gap of unknown length
53075	54010: contig of 936 bp in length
	gap of unknown length
54011	54906: contig of 896 bp in length
	gap of unknown length
54907	55797: contig of 891 bp in length
	gap of unknown length
55798	56687: contig of 890 bp in length
	gap of unknown length
56688	57579: contig of 892 bp in length
	gap of unknown length
57580	58443: contig of 864 bp in length
	gap of unknown length
58444	59345: contig of 902 bp in length
	gap of unknown length
59346	60251: contig of 906 bp in length
	gap of unknown length
60252	61197: contig of 946 bp in length
	gap of unknown length
61198	62066: contig of 869 bp in length
	gap of unknown length
62067	62960: contig of 894 bp in length
	gap of unknown length
62961	63871: contig of 911 bp in length

```
* * * gap of unknown length
* 63872 64769: contig of 898 bp in length
* * * gap of unknown length
* 64770 65714: contig of 945 bp in length
* * * gap of unknown length
* 65715 66596: contig of 882 bp in length
* * * gap of unknown length
* 66597 67499: contig of 903 bp in length
* * * gap of unknown length
* 67500 68408: contig of 909 bp in length
* * * gap of unknown length
* 68409 69331: contig of 923 bp in length
* * * gap of unknown length
* 69332 70264: contig of 933 bp in length

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
us-09-332-522b-2_copy_335_428 x AC011183 ..
Align seg 1/1 to: AC011183 from: 1 to: 72056

81 H15VALGUGlMetArgLysThr 88
|||||
23686 CATGTTGAGCAGATGAGGAAACT 23709

seq_name: gb_pr2:HS324L9
seq_documentation_block:
LOCUS HS324L9 73427 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from PAC 324L9 on chromosome 6p22.3-23. Contains
SYN.
ACCESSION AL023713 GI:3169110
VERSION AL023713.1
KEYWORDS 6p22.3-23.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 73427)
Mohammd, M.
Direct Submission
Submitted (06-JAN-1998) Chromosome 6 Project Group
(http://www.sanger.ac.uk/HGP/Chr6/) Sanger Centre, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humangenes@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
IMPORTANT: This sequence is the entire insert of clone 324L9.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variations annotated may not be found in the sequence submission
corresponding to the overlapping clone as we submit sequences with
only a small overlap as described above.
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed by the Sanger Centre chromosome 6
mapping group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6/
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key
The true left end of clone 324L9 is at 1 in this sequence. The true
right end of clone 324L9 is at 73427.
324L9 is from the library RPCI3 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong.
For further details see http://Pacpac.med.buffalo.edu/
Location/Qualifiers
1..73427
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p22.3-23"
/clone="RP3-324L9"
/clone_lib="RPCI-3"
1..670
/note="L1 repeat: matches 665..4 of consensus"
671..974
/note="Alu repeat: matches 297..1 of consensus"
3002..3899
/note="L1PA2 repeat: matches 891..10 of consensus"
3764..4272
/note="L1 repeat: matches 5390..4898 of consensus"
4270..8354
/note="L1 repeat: matches 4653..599 of consensus"
8454..8667
/note="L1 repeat: matches 566..381 of consensus"
8903..8995
/note="L1 repeat: matches 106..12 of consensus"
12433..12571
/note="MLTIB repeat: matches 251..390 of consensus"
16882..17792
/note="L1MA10 repeat: matches 153..1066 of consensus"
19387..19481
/note="L1ME1 repeat: matches 459..551 of consensus"
20210..20421
/note="MIR repeat: matches 11..261 of consensus"
20495..20522
/note="14 copies of 2 mer 100 & conserved"
20652..21534
/note="L1PA2 repeat: matches 891..1 of consensus"
21390..26791
/note="L1 repeat: matches 5390..4 of consensus"
27040..27141
/note="MIR repeat: matches 98..211 of consensus"
28188..28725
/note="MER48 repeat: matches 2..532 of consensus"
/complement(30121..30294)
/note="match: genomic DNA 277889"
31530..31692
/note="MIR repeat: matches 262..100 of consensus"
32111..32367
/note="Alu repeat: matches 46..301 of consensus;
incomplete repeat"
33142..33251
/note="L1R7 repeat: matches 1..131 of consensus"
33203..33514
/note="L1R7 repeat: matches 134..448 of consensus"
33914..34450
/note="L1 repeat: matches 4827..5390 of consensus"
34319..34612
/note="L1MC2 repeat: matches 1..314 of consensus"
34614..35075
/note="L1R7 repeat: matches 448..1 of consensus"
35073..35810
/note="L1MC3 repeat: matches 306..1114 of consensus"
36994..37374
/note="MS7A repeat: matches 426..1 of consensus"
37764..38332
/note="MER25 repeat: matches 1520..2104 of consensus"
38321..38567
/note="L1 repeat: matches 5141..5390 of consensus"
38422..39325
/note="L1PA13 repeat: matches 1..903 of consensus"
39885..40397
/note="L1PA16 repeat: matches 904..493 of consensus"
40424..40724
/note="AluB repeat: matches 302..3 of consensus"
40749..41429
/note="L1ME1 repeat: matches 221..915 of consensus"
42886..44757
/note="L1 repeat: matches 986..2855 of consensus"
```

```

repeat_region      44761..45120 repeat: matches 3..364 of consensus"
repeat_region      43121..44506 /note="HE1B-INTERNAL repeat: matches 1..397 of con
repeat_region      43501..45731 /note="HE1B-INTERNAL repeat: matches 431..660 of
consensus"
repeat_region      45730..46538 /note="HE1B-INTERNAL repeat: matches 767..1575 of
consensus"
repeat_region      46545..46906 /note="HE1C repeat: matches 1..371 of consensus"
repeat_region      46907..48475 /note="I1 repeat: matches 2852..4426 of consensus"
repeat_region      48475..49246 /note="I1 repeat: matches 4612..5390 of consensus"
repeat_region      49108..49316 /note="I1P2 repeat: matches 1..218 of consensus"
repeat_region      49324..49624 /note="I1P2 repeat: matches 1..218 of consensus"
repeat_region      49625..50309 /note="AluSg repeat: matches 303..1 of consensus"
repeat_region      50310..51615 /note="I1P13 repeat: matches 207..902 of consensus"
repeat_region      51616..51615 /note="AluSk repeat: matches 302..1 of consensus"
repeat_region      55687..56697 /note="AluSk repeat: matches 302..1 of consensus"
repeat_region      60783..61077 /note="MIR repeat: matches 75..190 of consensus"
repeat_region      62999..63295 /note="AluJ repeat: matches 302..12 of consensus"
repeat_region      63303..63564 /note="AluY repeat: matches 1..296 of consensus"
repeat_region      64425..64450 /note="MSTA repeat: matches 423..167 of consensus"
repeat_region      64575..64744 /note="I3 copies of 2 mer 96 % conserved"
repeat_region      65214..65386 /note="AluSc repeat: matches 130..299 of consensus;
incomplete repeat"
repeat_region      65460..65761 /note="MIR repeat: matches 250..63 of consensus"
repeat_region      67900..68132 /note="AluSp repeat: matches 303..1 of consensus"
repeat_region      70305..70421 /note="L1R10 repeat: matches 245..8 of consensus"
repeat_region      70989..71272 /note="MIR2 repeat: matches 22..146 of consensus"
repeat_region      72803..73101 /note="AluJ repeat: matches 282..1 of consensus;
incomplete repeat"
repeat_region      73128..73212 /note="AluSk repeat: matches 297..1 of consensus"
repeat_region      73128..73212 /note="MERSA repeat: matches 105..189 of consensus"
BASE COUNT      21620 a 14045 c 13920 g 23842 t
ORIGIN
alignment_scores:
      Quality:      8.00      Length:      8
      Ratio:      1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x HS324L9/rev ..
Align seg 1/1 to reverse of: HS324L9 from: 1 to: 73427

54 LysEuserLySSerAlaThrLeu 61
|||||
16615 AACTGTCCAATCTGCCATTGG 16592

```

ID	AC005395	standard: DNA; HTG: 83639 BP.
XX	AC	AC005395;
XX	SV	AC005395.8
DT	20-NOV-1998 (Rel. 57, Created)	
DT	03-SEP-1999 (Rel. 60, Last updated, Version 8)	
DE	Drosophila melanogaster chromosome 3 clone DS07437 (D469) map	
DE	8361-8362 strain y: cn bw sp. WORKING DRAFT SEQUENCE, 3 unordered	
DE	pieces.	
XX	HTG:	HTGS_PHASE1.
KM	Drosophila melanogaster (fruit fly)	
OS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota	
CC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;	
CC	Drosophilidae; Drosophila.	
XX	[1]	
RN	1-83639	
RP	Centner S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,	
RA	Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,	
RA	Fatfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,	
RA	Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,	
RA	Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,	
RA	Nixon K., Paaleb J.M., Park S., Pfeiffer B., Poon L., Sequiera A.,	
RA	Sehli H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,	
RA	Zieran L.L., Rubin G.M.;	
RT	"Sequencing of Drosophila melanogaster";	
RL	Unpublished.	
XX	[2]	
RN	1-83639	
RP	Centner S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,	
RA	Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,	
RA	Fatfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,	
RA	Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,	
RA	Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,	
RA	Nixon K., Paaleb J.M., Park S., Pfeiffer B., Poon L., Sequiera A.,	
RA	Sequeira A., Sehli H., Snir E., Svirskas R.R., Twomey B., Wan K.H.,	
RA	Weinburg T., Zhang R., Zieran L.L., Rubin G.M.;	
RT	Submitted (19-NOV-1998) to the EMBL/GenBank/DBJ databases.	
RL	Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121,	
RL	Berkeley, CA 94720, USA	
XX	On Apr 13, 1999 this sequence version replaced g1:4531419.	
CC	For further information about this sequence, including its location	
CC	and relationship to other sequences, please visit our sequence	
CC	archive Web site (http://www.fruitfly.org/sequence/) or send email	
CC	to dbd@fruitfly.berkeley.edu . All contigs in this submission meet	
CC	the following cutoffs: length >= 400 bases, pinrap computed error	
CC	rate <= 1/10. Pl library draft: sequence. It currently	
CC	* NOTE: This is a 'working draft' sequence. It consists of the pieces	
CC	* consists of 3 contigs. The true order of the pieces	
CC	* is not known and their order in this sequence record is	
CC	* arbitrary. Gaps between the contigs are represented as	
CC	* runs of N, but the exact sizes of the gaps are unknown.	
CC	* This record will be updated with the finished sequence	
CC	* as soon as it is available and the accession number will	
CC	* be preserved.	
CC	1	5896: contig of 5896 bp in length
CC	5897	5976: gap of unknown length
CC	5977	17680: contig of 11704 bp in length
CC	17681	17760: gap of unknown length
CC	17761	83639: contig of 65879 bp in length.
XX	Key	Location/Qualifiers
FH	source	1..83639
FT		/chromosome="3"


```

FT      /db_xref="taxon:7227"
FT      /organism="Drosophila melanogaster"
FT      /clone="DS07437 (D469)"
FT      /clone_lib="P1 library, partial Sau3A in pAD10sacBII"
FT      /map="83f1-83f2"
FT      /strain="y; cn bw sp"
XX
XX      Sequence 83639 BP; 23980 A; 17825 C; 17372 G; 24301 T; 161 other;

Alignment_scores:
      Quality:      8.00      Length:      8
      Ratio:      1.000      Gaps:      0
      Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522b-2_copy_335_428 x AC005985/rev ..
Align seg 1/1 to reverse of: AC005985 from: 1 to: 83639

54 LysleuserlysseralatHrleu 61
|||||
6421 AACGTCTCAAGTCGCGACTCTT 6398

seq_name: gb_in2:AC001652

seq_documentation_block:
LOCUS      AC001652      85757 bp      DNA      INV      17-JUL-1998
DEFINITION      Drosophila melanogaster DNA sequence (P1 DS00004 (D12)), complete
sequence.
ACCESSION      AC001652 L76986 L32616 L46822 L81326 L79833 L32614 L81325 L42048
L79834 L46820 L32619 L43416 L32618 L32615 L42053 L42049 L42051
L42050 L76988 L42052 L76987 L81327 L32621 L46821 L32612 L43415
L76985 L79835 L32622 L32613 L32617 L32620
AC001652.1 GI:2341048
HTG.
Drosophila melanogaster (Subclones in tet from P1 clone DS00004
(D12)) DNA.
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 85757)
AUTHORS      Gelniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
Sylizkas,R.R., Harris,N.L., Agbayan,A., Arcana,T.T., Baxter,E.,
Blazer,J.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,S.H., Lee,B., Lomoto,M.A., Mak,J., Mazda,P., Mok,M.S.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Punch,E., Sutr,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
Sequencing of antennapedia complex, homeotic genes
Unpublished (1997)
2 (bases 1 to 85757)
TITLE      Martin,C.H., Arcana,T.T., Bondoc,M.M., Chang,A., Critz,P.A.,
JOURNAL      Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E., Gunning,K.M.,
REFERENCE      Houston,K.A., Jaklevic,M.A., Kader,K.E., Kim,K., Kim,S.F.,
AUTHORS      Ko,C.L., Lewis,K.D., Li,M., Lindquist,K.J., Lomoto,M.A.,
Lustre,V.M., Machrus,M.U., Mayeda,C.A., Miguel,T.M., Miller,C.A.,
Mok,M.S., Pacleb,J.M., Patel,S.G., Santos,R.F., Subramanian,S.,
Wan,K.H., Whitelaw,K.R., Yee,A., Yen,R.T., Yu,C. and Palazzolo,M.J.
Direct Submission
Submitted (22-APR-1997)
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
for further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
email to drosophila@bgc.lbl.gov.
Library location: 4-1
Location/Qualifiers
FEATURES

```

```

source      1. 85757
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /chromosome="3R"
            /map="84A1-84A2"
            /clone="P1 DS00004 (D12)"
            /note="This sequence has not changed since its original
            submission on 04/22/1997. It was resubmitted in order to
            include all secondary accession numbers for the subclones
            belonging to this clone."

BASE COUNT      24009 a 18249 c 18762 g 24737 t
ORIGIN

Alignment_scores:
      Quality:      8.00      Length:      8
      Ratio:      1.000      Gaps:      0
      Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522b-2_copy_335_428 x AC001652 ..
Align seg 1/1 to: AC001652 from: 1 to: 85757

54 LysleuserlysseralatHrleu 61
|||||
14623 AACGTCTCAAGTCGCGACTCTT 14646

seq_name: gb_htg2:AL136180

seq_documentation_block:
LOCUS      AL136180      99487 bp      DNA      HTG      01-FEB-2000
DEFINITION      Homo sapiens chromosome 6 clone RP3-477B23, *** SEQUENCING IN
PROGRESS ***, 8 unordered pieces.
ACCESSION      AL136180
VERSION      AL136180.1 GI:6982309
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 99487)
AUTHORS      Sims,S.
JOURNAL      Direct Submission
Submitted (18-JAN-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Feb 16, 2000 this sequence version replaced gi:6689798.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
page etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 00017 Length: 1162bp
Contig_ID: 00127 Length: 13765bp
Contig_ID: 00138 Length: 1384bp
Contig_ID: 00709 Length: 42612bp
Contig_ID: 00828 Length: 1082bp
Contig_ID: 00836 Length: 1082bp
Contig_ID: 00896 Length: 5150bp
Contig_ID: 00906 Length: 1467bp
Contig_ID: 00963 Length: 27265bp.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved. 1162: contig of 1162 bp in length
* 1
* 1163 1962: gap of 800 bp
* 15737: contig of 13765 bp in length

```

* 15728 16527: gap of 800 bp
* 16528 17911: contig of 1384 bp in length
* 17912 18711: gap of 800 bp
* 18712 61323: contig of 42612 bp in length
* 61324 62123: gap of 800 bp
* 62124 63205: contig of 1082 bp in length
* 63206 64005: gap of 800 bp
* 64006 69155: contig of 5150 bp in length
* 69156 69955: gap of 800 bp
* 69956 71422: contig of 1467 bp in length
* 71423 72223: gap of 800 bp
* 72223 99487: contig of 27265 bp in length.
Location/Qualifiers
1..99487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP3-477B23"
/clone_1lb="RPCI-3"
BASE COUNT 26800 a 18017 c 17863 g 31206 t 5601 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x AL137159 ..
Align seg 1/1 to: AL136180 from: 1 to: 99487

54 LysleuSerIysSerAlaThrIeu 61
|||||||
63131 AACGTGTCACAAATCGCACCTTG 63154

seq_name: gb_htg2:AL137159

seq_documentation_block:
LOCUS AL137159 107860 bp DNA HTG 01-FEB-2000
DEFINITION Homo sapiens chromosome 1 clone RP4-713B5 map p21.2-22.2, ***
SEQUENCING IN PROGRESS ***, 3 unordered pieces.
ACCESSION AL137159
VERSION AL137159.1 GI:6982380
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 107860)
Pavitt,R.
Direct Submission
Submitted (24-JAN-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Feb 16, 2000 this sequence version replaced g1:6752234.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 00659 Length: 76901bp
Contig_ID: 00855 Length: 1399bp
Contig_ID: 00881 Length: 1399bp
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

* be preserved.
* 1 76901: contig of 76901 bp in length
* 76902 77701: gap of 800 bp
* 77702 105661: contig of 27960 bp in length
* 105662 106461: gap of 800 bp
* 106462 107860: contig of 1399 bp in length.
Location/Qualifiers
1..107860
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p21.2-22.2"
/clone="RP4-713B5"
/clone_1lb="RPCI-4"
BASE COUNT 32644 a 21036 c 20050 g 32530 t 1600 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x AL137159/rev ..
Align seg 1/1 to reverse of: AL137159 from: 1 to: 107860

82 ValGIuGlnMetArgIysThrIeu 89
|||||||
27998 GTTGAAACAAATCGGAAACTCTA 27975

seq_name: gb_htg2:HSDJ399E4

seq_documentation_block:
LOCUS HSDJ399E4 110771 bp DNA HTG 25-FEB-2000
DEFINITION Homo sapiens chromosome 6 clone RP3-399E4, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL121948
VERSION AL121948.7 GI:7106641
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 110771)
Parker,A.
Direct Submission
Submitted (25-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Feb 27, 2000 this sequence version replaced g1:6982797.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 00741 Length: 110771bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
Location/Qualifiers
1..110771
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP3-399E4"
/clone_1lb="RPCI-3"
BASE COUNT 33581 a 22908 c 21979 g 32302 t 1 others
ORIGIN

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

alignment_scores: Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522b-2_COPY_335_428 x HSDJ399E4 ..

Align seg 1/1 to: HSDJ399E4 from: 1 to: 110771

58 SerAlaThrLeuArgAlaIle 65
|||||
67766 TCCGCCACCTGAGAGAGCAATC 67789

seq_name: gb_htg5:AC009557

seq_documentation_block:
LOCUS AC009557 128590 bp DNA HTG 25-OCT-1999
DEFINITION Homo sapiens chromosome 15 clone 76_D_16 map 15, LOW-PASS SEQUENCE.
SAMPLING.
ACCESSION AC009557
VERSION AC009557.2 GI:5114932
KEYWORDS HTG; HTGS_PHRASEO.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 128590)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Castle,A., Cerny,J., Collangelo,M., Collins,S., Collins,A.,
 Cooke,P., Dearlano,K., Depayre,E., Devon,K., Dewar,K.,
 Dorelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
 Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
 Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
 Karats,A., Lehoczy,J., Liu,C., Locke,K., MacDonald,P.,
 Margulis,N., McEwan,P., McGurk,A., McKernan,K., McDonald,J.,
 Meidum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
 Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
 Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
 Teefaye,S., Torrella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
 Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
 Direct Submission
COMMENT Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Oct 25, 1999 this sequence version replaced g1:5788087.
 All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html.
 * NOTE: This record contains 157 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 749: contig of 749 bp in length
* 1 749: contig of 749 bp in length
* 750 1492: contig of 743 bp in length
* 750 1492: contig of 743 bp in length
* 1493 2232: contig of 740 bp in length
* 1493 2232: contig of 740 bp in length
* 2233 2976: contig of 744 bp in length
* 2233 2976: contig of 744 bp in length

* 2977 3733: gap of unknown length
* 2977 3733: contig of 757 bp in length
* 3734 4520: gap of unknown length
* 3734 4520: contig of 787 bp in length
* 4521 5270: gap of unknown length
* 4521 5270: contig of 750 bp in length
* 5271 6042: gap of unknown length
* 5271 6042: contig of 772 bp in length
* 6043 6807: gap of unknown length
* 6043 6807: contig of 765 bp in length
* 6808 7568: gap of unknown length
* 6808 7568: contig of 761 bp in length
* 7569 8295: gap of unknown length
* 7569 8295: contig of 727 bp in length
* 8296 9028: gap of unknown length
* 8296 9028: contig of 733 bp in length
* 9029 9765: gap of unknown length
* 9029 9765: contig of 737 bp in length
* 9766 10494: gap of unknown length
* 9766 10494: contig of 729 bp in length
* 10495 11223: gap of unknown length
* 10495 11223: contig of 729 bp in length
* 11224 11996: gap of unknown length
* 11224 11996: contig of 773 bp in length
* 11997 12768: gap of unknown length
* 11997 12768: contig of 772 bp in length
* 12769 13540: gap of unknown length
* 12769 13540: contig of 772 bp in length
* 13541 14293: gap of unknown length
* 13541 14293: contig of 753 bp in length
* 14294 15045: gap of unknown length
* 14294 15045: contig of 752 bp in length
* 15046 15801: gap of unknown length
* 15046 15801: contig of 756 bp in length
* 15802 16566: gap of unknown length
* 15802 16566: contig of 765 bp in length
* 16567 17307: gap of unknown length
* 16567 17307: contig of 741 bp in length
* 17308 18075: gap of unknown length
* 17308 18075: contig of 768 bp in length
* 18076 18805: gap of unknown length
* 18076 18805: contig of 730 bp in length
* 18806 19518: gap of unknown length
* 18806 19518: contig of 713 bp in length
* 19519 20239: gap of unknown length
* 19519 20239: contig of 721 bp in length
* 20240 20937: gap of unknown length
* 20240 20937: contig of 698 bp in length
* 20938 21679: gap of unknown length
* 20938 21679: contig of 742 bp in length
* 21680 22449: gap of unknown length
* 21680 22449: contig of 770 bp in length
* 22450 23206: gap of unknown length
* 22450 23206: contig of 757 bp in length
* 23207 23957: gap of unknown length
* 23207 23957: contig of 751 bp in length
* 23958 24714: gap of unknown length
* 23958 24714: contig of 757 bp in length
* 24715 25475: gap of unknown length
* 24715 25475: contig of 761 bp in length
* 25476 26219: gap of unknown length
* 25476 26219: contig of 744 bp in length
* 26220 26966: gap of unknown length
* 26220 26966: contig of 747 bp in length
* 26967 27698: gap of unknown length
* 26967 27698: contig of 732 bp in length
* 27699 28414: gap of unknown length
* 27699 28414: contig of 716 bp in length
* 28415 29142: gap of unknown length
* 28415 29142: contig of 728 bp in length
* 29143 29902: gap of unknown length
* 29143 29902: contig of 760 bp in length

```

* 29903 30671: contig of 769 bp in length
*      gap of unknown length
* 30672 31436: contig of 765 bp in length
*      gap of unknown length
* 31437 32184: contig of 748 bp in length
*      gap of unknown length
* 32185 32924: contig of 740 bp in length
*      gap of unknown length
* 32925 33664: contig of 740 bp in length
*      gap of unknown length
* 33665 34376: contig of 712 bp in length
*      gap of unknown length
* 34377 35120: contig of 744 bp in length
*      gap of unknown length
* 35121 35838: contig of 718 bp in length
*      gap of unknown length
* 35839 36578: contig of 740 bp in length
*      gap of unknown length
* 36579 37306: contig of 728 bp in length
*      gap of unknown length
* 37307 38050: contig of 744 bp in length
*      gap of unknown length
* 38051 38823: contig of 773 bp in length
*      gap of unknown length
* 38824 39598: contig of 775 bp in length
*      gap of unknown length
* 39599 40345: contig of 747 bp in length
*      gap of unknown length
* 40346 41062: contig of 717 bp in length
*      gap of unknown length
* 41063 41833: contig of 771 bp in length
*      gap of unknown length
* 41834 42576: contig of 743 bp in length
*      gap of unknown length
* 42577 43331: contig of 755 bp in length
*      gap of unknown length
* 43332 44065: contig of 734 bp in length
*      gap of unknown length
* 44066 44787: contig of 722 bp in length
*      gap of unknown length
* 44788 45518: contig of 731 bp in length
*      gap of unknown length
* 45519 46261: contig of 743 bp in length
*      gap of unknown length
* 46262 47038: contig of 777 bp in length
*      gap of unknown length
* 47039 47787: contig of 749 bp in length
*      gap of unknown length
* 47788 48559: contig of 772 bp in length
*      gap of unknown length
* 48560 49318: contig of 759 bp in length
*      gap of unknown length
* 49319 50055: contig of 737 bp in length
*      gap of unknown length
* 50056 50790: contig of 735 bp in length
*      gap of unknown length
* 50791 51510: contig of 720 bp in length
*      gap of unknown length
* 51511 52241: contig of 731 bp in length
*      gap of unknown length
* 52242 52972: contig of 731 bp in length
*      gap of unknown length
* 52973 53690: contig of 718 bp in length
*      gap of unknown length
* 53691 54449: contig of 759 bp in length
*      gap of unknown length
* 54450 55215: contig of 766 bp in length
*      gap of unknown length
* 55216 55954: contig of 739 bp in length
*      gap of unknown length
* 55955 56675: contig of 721 bp in length
*      gap of unknown length

```

```

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-09-332-522b-2_COPY_335_428 x AC009557
  Align seq 1/1 to: AC009557 from: 1 to: 128590

42 GlnGlnLeuYsValLeuLeuCs 49
|||||
71750 CAGCACTCAAGTATATGTGT 71773

seq_name: gb_hvg2:AL157902

seq_documentation_block:
LOCUS      AL157902 142249 bp      DNA      HTG      24-FEB-2000
DEFINITION Homo sapiens chromosome 1 clone RP4-675G20 map p13.2, ***
SEQUENCING IN PROGRESS ***, 16 unordered pieces.
ACCESSION  AL157902
VERSION    AL157902.1 GI:7105850
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 142249)
REFERENCE  Pavlitt, R.
            Direct SubMISSION
            Submitted (24-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: CloneRequests@sanger.ac.uk
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in
            progress and the release of this data is based on the understanding
            that the sequence may change as work continues. The sequence may
            be contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Contig_ID: 00083 Length: 9095bp
            Contig_ID: 00198 Length: 1378bp
            Contig_ID: 00261 Length: 1649bp
            Contig_ID: 00381 Length: 1053bp
            Contig_ID: 00408 Length: 1092bp
            Contig_ID: 00462 Length: 1065bp
            Contig_ID: 00496 Length: 1948bp
            Contig_ID: 00505 Length: 9137bp
            Contig_ID: 00572 Length: 1067bp
            Contig_ID: 00638 Length: 1280bp
            Contig_ID: 00674 Length: 1039bp
            Contig_ID: 01255 Length: 1226bp
            Contig_ID: 01385 Length: 1061bp
            Contig_ID: 01426 Length: 84025bp
            Contig_ID: 01573 Length: 1200bp
            Contig_ID: 01602 Length: 3034bp.
            NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            1 9095: contig of 9095 bp in length
            * 9096 9895: gap of 800 bp
            * 9896 11273: contig of 1378 bp in length
            * 11274 12073: gap of 800 bp
            * 12074 13722: contig of 1649 bp in length
            * 13723 14522: gap of 800 bp
            * 14523 15575: contig of 1053 bp in length
            * 15576 16375: gap of 800 bp
            * 16376 27367: contig of 10992 bp in length
            * 27368 28167: gap of 800 bp

```

* 28168 29232: contig of 1065 bp in length
* 29233 30032: gap of 800 bp
* 30033 31980: contig of 1948 bp in length
* 31981 32780: gap of 800 bp
* 32781 41917: contig of 9137 bp in length
* 41918 42717: gap of 800 bp
* 42718 43784: contig of 1067 bp in length
* 43785 44584: gap of 800 bp
* 44585 45864: contig of 1280 bp in length
* 45865 46664: gap of 800 bp
* 46665 47703: contig of 1039 bp in length
* 47704 48503: gap of 800 bp
* 48504 49730: contig of 1226 bp in length
* 49730 50529: gap of 800 bp
* 50530 51590: contig of 1061 bp in length
* 51591 52390: gap of 800 bp
* 52391 136415: contig of 84025 bp in length
* 136416 137215: gap of 800 bp in length
* 137216 138415: contig of 1200 bp in length
* 138416 139215: gap of 800 bp
* 139216 142249: contig of 3034 bp in length.
Location/Qualifiers
1. 142249
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p13.2"
/clone="RP4-675C20"
/clone_11b="RRC1-4"
BASE COUNT 38253 a 26772 c 27862 g 37347 t 12015 others
ORIGIN
Alignment scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
Alignment block:
US-09-332-522b-2_COPY_335_428 x AL157902 ..
Align seg 1/1 to: AL157902 from: 1 to: 142249
46 ValLeuLeuCyGgLyAspGluAla 53
|||||
38614 GTTTRACTGTGCGTGCAGACGA 38637
seq_name: gb_htg2:AL136133
seq_documentation_block:
LOCUS AL136133 142394 bp DNA HTG 03-FEB-2000
DEFINITION Homo sapiens chromosome 1 clone CTA-253M17, *** SEQUENCING IN
PROGRESS ***, 31 unordered pieces.
ACCESSION AL136133
VERSION AL136133.1 GI:6982515
KEYWORDS HTG; HTGS_PHASEL; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 142394)
Pavitt,R.
Direct Submission
Submitted (03-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requests: clonerequests@sanger.ac.uk
On Feb 16, 2000 this sequence version replaced gi:6686696.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate

segments. Contig_ID: 00018 Length: 3722bp
Contig_ID: 00071 Length: 4523bp
Contig_ID: 00094 Length: 5769bp
Contig_ID: 00133 Length: 1508bp
Contig_ID: 00143 Length: 2462bp
Contig_ID: 00156 Length: 1387bp
Contig_ID: 00163 Length: 2883bp
Contig_ID: 00173 Length: 2337bp
Contig_ID: 00178 Length: 3715bp
Contig_ID: 00215 Length: 1897bp
Contig_ID: 00251 Length: 4122bp
Contig_ID: 00266 Length: 1185bp
Contig_ID: 00267 Length: 1185bp
Contig_ID: 00272 Length: 3770bp
Contig_ID: 00280 Length: 13393bp
Contig_ID: 00286 Length: 1142bp
Contig_ID: 00388 Length: 8316bp
Contig_ID: 00403 Length: 1990bp
Contig_ID: 00413 Length: 1711bp
Contig_ID: 00468 Length: 3304bp
Contig_ID: 00529 Length: 4230bp
Contig_ID: 00550 Length: 4604bp
Contig_ID: 00571 Length: 6764bp
Contig_ID: 00571 Length: 1018bp
Contig_ID: 00630 Length: 6972bp
Contig_ID: 00690 Length: 4333bp
Contig_ID: 00704 Length: 1727bp
Contig_ID: 00857 Length: 3649bp
Contig_ID: 00874 Length: 8723bp
Contig_ID: 00880 Length: 4723bp
Contig_ID: 00890 Length: 1067bp.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known, and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 3722: contig of 3722 bp in length
* 3723 4522: gap of 800 bp
* 4523 9045: contig of 4523 bp in length
* 9046 9845: gap of 800 bp
* 9846 15614: contig of 5769 bp in length
* 15615 16414: gap of 800 bp
* 16415 18323: contig of 1908 bp in length
* 18323 19122: gap of 800 bp
* 19123 21584: contig of 2462 bp in length
* 21585 22384: gap of 800 bp
* 22385 23771: contig of 1387 bp in length
* 23772 24571: gap of 800 bp
* 24572 27454: contig of 2883 bp in length
* 27455 28254: gap of 800 bp
* 28255 30591: contig of 2337 bp in length
* 30592 31391: gap of 800 bp
* 31392 35106: contig of 3715 bp in length
* 35107 35906: gap of 800 bp
* 35907 37803: contig of 1897 bp in length
* 37804 38603: gap of 800 bp
* 38604 42725: contig of 4122 bp in length
* 42726 43525: gap of 800 bp
* 43526 44710: contig of 1185 bp in length
* 44711 45510: gap of 800 bp
* 45511 46665: contig of 1155 bp in length
* 46666 47465: gap of 800 bp
* 47466 51235: contig of 3770 bp in length
* 51236 52035: gap of 800 bp
* 52036 65328: contig of 13293 bp in length
* 65329 66128: gap of 800 bp
* 66130 67270: contig of 1142 bp in length
* 67271 68070: gap of 800 bp
* 68071 76386: contig of 8316 bp in length
* 76387 77186: gap of 800 bp

77187 79176: contig of 1990 bp in length
* 79177 79976: gap of 800 bp in length
* 79977 81687: contig of 1711 bp in length
* 81688 82487: gap of 800 bp
* 82488 85791: contig of 3304 bp in length
* 85792 86591: gap of 800 bp
* 86592 90821: contig of 4230 bp in length
* 90822 91621: gap of 800 bp in length
* 91622 96225: contig of 4604 bp in length
* 96226 97025: gap of 800 bp
* 97026 103789: contig of 6764 bp in length
* 103790 104589: gap of 800 bp
* 104590 105607: contig of 1018 bp in length
* 105608 106407: gap of 800 bp in length
* 106408 113379: contig of 6972 bp in length
* 113380 114179: gap of 800 bp
* 114180 118502: contig of 4323 bp in length
* 118503 119302: gap of 800 bp
* 119303 121029: contig of 1727 bp in length
* 121030 121829: gap of 800 bp
* 121830 125478: contig of 3649 bp in length
* 125479 126278: gap of 800 bp
* 126279 135004: contig of 8726 bp in length
* 135005 135804: gap of 800 bp
* 135805 140527: contig of 4723 bp in length
* 140528 141327: gap of 800 bp
* 141328 142394: contig of 1067 bp in length.
Location/Qualifiers
1. 142394
/organism="Homo sapiens"
/chromosome="1"
/clone="CTA-253N17"
/clone_lib="CT1978SK-A1"
BASE COUNT 36932 a 22482 c 21772 g 37205 t 24003 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x AL136133 ..
Align seg 1/1 to: AL136133 from: 1 to: 142394

41 ||le|g|n|l|e|u|l|y|v|a|l|l|e|u|l|e|u| 48
|||||||
60696 ATTCACAGCTGAGGTATTATTA 60719

seq_name: gb_htg12:AC011350

seq_documentation_block:
LOCUS AC011350 148298 bp DNA HTG 18-FEB-2000
DEFINITION Homo sapiens chromosome 5 clone CTC-317E24, WORKING DRAFT SEQUENCE,
20 unordered pieces.
AC011350
VERSION AC011350.3 GI:6997087
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 148298)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 148298)
DOE Joint Genome Institute.
Direct Submission
Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT On Feb 18, 2000 this sequence version replaced gi:6604446.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>
-----Summary Statistics
Consensus quality: 126262 bases at least Q40
Consensus quality: 136960 bases at least Q30
Consensus quality: 141075 bases at least Q20
Estimated insert size: 148298; sum-of-contigs estimation
Estimated insert size: 148000; pulse field gel estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1463 2892: contig of 1430 bp in length
2893 4542: gap of 1650 bp in length
4543 5685: gap of unknown length
5686 7689: gap of unknown length
7690 10844: contig of 3155 bp in length
10845 14290: contig of 3446 bp in length
14291 15575: gap of unknown length
15576 17385: contig of 1810 bp in length
17386 23911: contig of 6526 bp in length
23912 27515: contig of 3604 bp in length
27516 30837: contig of 3322 bp in length
30838 38790: contig of 7953 bp in length
38791 44392: contig of 5602 bp in length
44393 58718: contig of 14326 bp in length
58719 65296: contig of 6578 bp in length
65297 81376: contig of 16080 bp in length
81377 99997: contig of 18621 bp in length
99998 119925: contig of 19928 bp in length
119926 148298: contig of 28373 bp in length.
Location/Qualifiers
1. 148298
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-317E24"
BASE COUNT 43740 a 27637 c 28259 g 48598 t 64 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000


```

REFERENCE      1 (bases 1 to 176932)
AUTHORS        Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE          Palmer's.
JOURNAL        Submitted (28-APR-1997) Chromosome 6 Project Group
                (http://www.sanger.ac.uk/chrf6/); Sanger Centre, Hinxton,
                Cambridgeshire, CB10 1RO, UK. E-mail enquiries:
                humphrey@sanger.ac.uk; clone requests: clonerequest@sanger.ac.uk
                On Apr 30, 1997 this sequence version replaced g1:1848177.
                IMPORTANT: This sequence is the entire insert of clone 121G13. The
                true left end of clone 121G13 is at 1 in this sequence. The true
                right end of clone 121G13 is at 176932.
                121G13 is from the human PAC library described in Ioannou A.P. et
                al Nature Genet 6, 84-89.
                de Jong P.J., enquiries: http://bacpac.med.buffalo.edu/.
                Location/Qualifiers
FEATURES       1..176932
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="6"
                /map="6q16"
                /clone_lib="RPC1-1"
                /clone="RP1-121G13"
                192..1310
                /note="MER42c repeat: matches 400..1536 of consensus"
                repeat_region
                2167..2466
                /note="Alu1 repeat: matches 301..1 of consensus"
                repeat_region
                6275..6785
                /note="L1MA9 repeat: matches 1048..506 of consensus"
                repeat_region
                6791..7475
                /note="L1MA9 repeat: matches 680..1 of consensus"
                repeat_region
                7290..8073
                /note="L1 repeat: matches 5389..4612 of consensus"
                repeat_region
                8081..8694
                /note="L1 repeat: matches 4071..3443 of consensus"
                repeat_region
                8711..8820
                /note="L1P4 repeat: matches 886..777 of consensus"
                repeat_region
                8820..9589
                /note="L1 repeat: matches 3455..2668 of consensus"
                repeat_region
                11846..12147
                /note="AluX repeat: matches 1..302 of consensus"
                repeat_region
                12148..12375
                /note="57 copies of 4 mer 82 & conserved"
                repeat_region
                12376..12415
                /note="10 copies of 4 mer 88 & conserved"
                repeat_region
                17683..17965
                /note="Alu1 repeat: matches 34..300 of consensus;
                incomplete repeat"
                repeat_region
                17967..18002
                /note="18 copies of 2 mer 81 & conserved"
                repeat_region
                18477..18774
                /note="AluX repeat: matches 1..300 of consensus"
                repeat_region
                19234..19531
                /note="AluX repeat: matches 301..1 of consensus"
                repeat_region
                21080..21226
                /note="Mir2 repeat: matches 145..1 of consensus"
                repeat_region
                21802..22093
                /note="AluX repeat: matches 294..1 of consensus"
                repeat_region
                22237..22419
                /note="L1MA9 repeat: matches 857..1032 of consensus"
                repeat_region
                22937..23042
                /note="Mir2 repeat: matches 142..36 of consensus"
                repeat_region
                24027..24321
                /note="Alu1 repeat: matches 300..11 of consensus"
                repeat_region
                24341..24413
                /note="Mir repeat: matches 177..249 of consensus"
                repeat_region
                24551..24904
                /note="TIRIA repeat: matches 1..354 of consensus"
                repeat_region
                28247..28607
                /note="MER44A repeat: matches 1..330 of consensus"
                repeat_region
                28678..28974
                /note="AluX repeat: matches 1..297 of consensus"
                repeat_region
                31497..31755
                /note="Alu1 repeat: matches 45..301 of consensus;
                incomplete repeat"
                repeat_region
                31819..32000
                /note="MER5A repeat: matches 189..4 of consensus"
                repeat_region
                33167..33320
                /note="Alu1 repeat: matches 125..276 of consensus;
                incomplete repeat"
                repeat_region
                33322..33357
                /note="18 copies of 2 mer 89 & conserved"
                repeat_region
                33405..33504
                /note="Mir2 repeat: matches 37..146 of consensus"
                repeat_region
                34977..35068
                /note="Mir repeat: matches 28..121 of consensus"
                repeat_region
                35386..35613
                /note="14 copies of 2 mer 96 & conserved"
                repeat_region
                35871..35951
                /note="Mir2 repeat: matches 59..139 of consensus"
                repeat_region
                36155..36281
                /note="Mir repeat: matches 260..131 of consensus"
                repeat_region
                36334..36626
                /note="AluX repeat: matches 294..1 of consensus"
                repeat_region
                36627..36767
                /note="Mir repeat: matches 142..5 of consensus"
                repeat_region
                37519..37963
                /note="L1 repeat: matches 4943..5390 of consensus"
                repeat_region
                37817..38699
                /note="L1P2 repeat: matches 3..899 of consensus"
                repeat_region
                39703..39943
                /note="Mir repeat: matches 262..20 of consensus"
                repeat_region
                44095..44610
                /note="match: 279048 flow sorted chromosome 6 fragment;
                match: M1135 M11354 Histone H3"
                misc_feature
                47661..48117
                /note="match: 256607 256608; Cpg island genomic MseI
                fragment, clone 12e4"
                repeat_region
                49482..49541
                /note="20 copies of 3 mer 82 & conserved"
                repeat_region
                50288..50323
                /note="9 copies of 4 mer 100 & conserved"
                repeat_region
                57579..57662
                /note="42 copies of 2 mer 95 & conserved"
                repeat_region
                63787..65052
                /note="Putative Cpg Island"
                misc_feature
                66644..66944
                /note="match: 263936; Cpg Island genomic MseI fragment,
                clone 93a12"
                repeat_region
                68327..68354
                /note="7 copies of 4 mer 93 & conserved"
                repeat_region
                69340..69524
                /note="MER5A repeat: matches 187..1 of consensus"
                repeat_region
                70893..71341
                /note="L1MD2 repeat: matches 573..89 of consensus"
                repeat_region
                71869..71958
                /note="3 copies of 30 mer 81 & conserved"
                repeat_region
                74044..74142
                /note="Mir repeat: matches 181..85 of consensus"
                repeat_region
                74321..74616
                /note="Alu1 repeat: matches 300..5 of consensus"
                repeat_region
                74625..75495
                /note="L1MD2 repeat: matches 1076..244 of consensus"
                repeat_region
                75496..75773
                /note="Alu1 repeat: matches 302..1 of consensus"
                repeat_region
                75774..76006
                /note="L1MD2 repeat: matches 247..10 of consensus"
                repeat_region
                75875..76132
                /note="L1 repeat: matches 5379..5111 of consensus"
                repeat_region
                77045..77350
                /note="Alu1 repeat: matches 3..301 of consensus"
                repeat_region
                77429..77671
                /note="L1MD3 repeat: matches 304..542 of consensus"
                repeat_region
                78010..78301
                /note="AluX repeat: matches 2..303 of consensus"
                repeat_region
                78331..78469

```



```

/note="Aluud repeat: matches 273. .135 of consensus;
incomplete repeat"
repeat_region 79831..79977
/note="MIR repeat: matches 262. .93 of consensus"
80190..80477
/note="Aluud repeat: matches 296. .1 of consensus"
81341..81480
/note="MIR repeat: matches 15. .144 of consensus"
82971..83270
/note="Aluud repeat: matches 300. .1 of consensus"
89580..89699
/note="LTR1C repeat: matches 466. .348 of consensus"
89728..90252
/note="MER1A repeat: matches 1. .527 of consensus"
90258..90584
/note="LTR1C repeat: matches 326. .1 of consensus"
90342..90584
/note="LTR1B repeat: matches 250. .1 of consensus"
90614..90908
/note="AluSg repeat: matches 295. .2 of consensus"
93666..93837
/note="LTR1F repeat: matches 539. .342 of consensus"
93915..94151
/note="LTR1F repeat: matches 220. .1 of consensus"
95016..95408
/note="THE1C repeat: matches 371. .1 of consensus"
95494..96364
/note="LIMB3 repeat: matches 914. .2 of consensus"
96329..96530
/note="L1 repeat: matches 5287. .5075 of consensus"
96702..97327
/note="L1PAB repeat: matches 910. .267 of consensus"
97328..97625
/note="AluY repeat: matches 1. .297 of consensus"
97626..97898
/note="L1PAB repeat: matches 274. .1 of consensus"
97750..98663
/note="L1 repeat: matches 5389. .4454 of consensus"
98662..100026
/note="L1 repeat: matches 1473. .2831 of consensus"
101719..101766
/note="MIR2 repeat: matches 142. .94 of consensus"
101972..102155
/note="Aluud repeat: matches 120. .302 of consensus;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522b-2_COPY_335_428 x HS121G13 ..
Align seg 1/1 to: HS121G13 from: 1 to: 176932
58 SerAlaThLeuArGArGAlaIaIe 65
|||||
47835 TCCGCCACCTCGAGAGAGCAATC 47858

seq_name: gb_pr4:AC007406
seq_documentation_block:
LOCUS AC007406 177402 bp DNA PRI 18-JUN-1999
DEFINITION Homo sapiens 12p13.3 BAC RPII1-28313 (Roswell Park Cancer
Institute Human BAC Library) complete sequence.
ACCESSION AC007406
VERSION AC007406.1 GI:4689442
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 177402)

```

```

AUTHORS
Muzny,D., Aronson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,
Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H.,
Gorell,L.L., Hernandez,J., Isser,A., Jackson,L., Kneit,S.,
Kondejewski,N., Lau,S., Leal,B., Lee,E., Licharge,O., Liu,W.,
Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A.,
Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,
Rashid,N.D., Raves,C., Scherer,S.E., Shen,H., Shim,C., Simon,M.,
Vo,Q., Williamson,A., Worley,K.C., Zhang,A.M., Yang,R., Yu,W.,
Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.

TITLE
Direct Submission
Unpublished
2 (bases 1 to 177402)

REFERENCE
Worley,K.C.
Direct Submission
Submitted (05-MAY-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 177402)

REFERENCE
Worley,K.C.
Direct Submission
Submitted (05-MAY-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 177402)

REFERENCE
Worley,K.C.
Direct Submission
Submitted (18-JUN-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

COMMENT
CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
html.

QUALSTAT-REPORT.
Location/Qualifiers
1..177402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RPII1-28313"
/chromosome="12p13.3"
FEATURES
source

```

```

misc_feature      3969..24129
                  /note="variable number of TCCA repeats"
                  /function="Ambiguities"
misc_feature      4834..5383
                  /note="Region: Similar to Unigene cluster containing
                  AT268554 and AA056734"
repeat_region    5030..5055
                  /rpt_family="CCTG)n"
STS              5228..5339
                  /standard_name="G30350"
                  /db_xref="dbSTS:44531"
                  /db_xref="dbSTS:35176"
repeat_region    complement(6073..6128)
                  /rpt_family="L2"
repeat_region    complement(7046..7108)
                  /rpt_family="L2"
repeat_region    complement(7809..8024)
                  /rpt_family="Alu"
repeat_region    8025..8249
                  /rpt_family="AluSg/x"
                  /rpt_family="AluSg/x"
                  /rpt_family="AluSg/x"
repeat_region    complement(8250..8363)
                  /rpt_family="Alu"
repeat_region    complement(8366..8459)
                  /rpt_family="Alu"
repeat_region    9796..10214
                  /rpt_family="LIM4"
                  /rpt_family="LIM4"
repeat_region    10218..10504
                  /rpt_family="Alu"
repeat_region    10508..10640
                  /rpt_family="Alu"
repeat_region    10649..10702
                  /rpt_family="Alu"
repeat_region    10766..11265
                  /rpt_family="LIM4"
                  /rpt_family="LIM4"
repeat_region    11266..11400
                  /rpt_family="Alu"
repeat_region    11401..11699
                  /rpt_family="Alu"
repeat_region    11700..11868
                  /rpt_family="Alu"
repeat_region    11869..12073
                  /rpt_family="LIM4"
                  /rpt_family="LIM4"
repeat_region    complement(13308..13591)
                  /rpt_family="MLT1F"
repeat_region    complement(14146..15546)
                  /rpt_family="PTRS"
misc_feature     14547..14627
                  /note="poly A and CA repeats, variable number of As"
                  /function="Ambiguities"
repeat_region    15583..15659
                  /rpt_family="CA)n"
gene             complement(17009..18134,19401..19571,20200..20300,
                  22148..22250,23047..23184,23693..23805,24300..24424,
                  24823..24926,25720..25854,27574..27706,28685..28772,
                  29663..29803,31487..31621,36696..36966,37128..37491,
                  38949..39033,40213..40249))
                  /gene="Human pephbct-1 betaine-GABA transporter mRNA
                  U27699"
STS              17022..17147
                  /standard_name="G21338"
                  /db_xref="dbSTS:35176"
repeat_region    17120..17184
                  /rpt_family="MIR"
repeat_region    18326..18491
                  /rpt_family="MIR"
repeat_region    18453..18496
                  /rpt_family="L2"
repeat_region    18498..18531
                  /rpt_family="L2"
repeat_region    18587..18606
                  /rpt_family="(TTA)n"
repeat_region    18614..18678
                  /rpt_family="(TGG)n"
repeat_region    21379..21472
                  /rpt_family="MLT1F"

```

```

repeat_region    /rpt_family="MIR"
repeat_region    22566..22868
                  /rpt_family="MLT1R"
repeat_region    23968..24131
                  /rpt_family="(TCCA)n"
repeat_region    24447..24667
                  /rpt_family="GA-rich"
repeat_region    25357..25524
                  /rpt_family="Alu"
repeat_region    26067..26107
                  /rpt_family="GA-rich"
repeat_region    complement(26151..26257)
                  /rpt_family="LIM4"
repeat_region    complement(27027..27329)
                  /rpt_family="Alu"
repeat_region    28270..28416
                  /rpt_family="MER63A"
repeat_region    29115..29412
                  /rpt_family="L2"
repeat_region    complement(29978..30260)
                  /rpt_family="Alu"
repeat_region    30408..30435
                  /rpt_family="GC-rich"
repeat_region    30714..30747
                  /rpt_family="GC-rich"
repeat_region    complement(32319..32544)
                  /rpt_family="MIR"
repeat_region    complement(32994..33438)
                  /rpt_family="LIM3"
STS              34381..34639
                  /standard_name="G41013"
                  /db_xref="dbSTS:61057"
                  /db_xref="dbSTS:61057"
repeat_region    34489..34512
                  /rpt_family="AT-rich"
repeat_region    35501..35756
                  /rpt_family="LIM8"
repeat_region    35856..36163
                  /rpt_family="L2"
repeat_region    36336..36394
                  /rpt_family="L2"
repeat_region    complement(36480..36541)
                  /rpt_family="MIR"
repeat_region    complement(37129..37234)

alignment_scores:
      Quality: 8.00      Length: 8
      Ratio: 1.000      Gaps: 0
      Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-332-522b-2_COPY_335_428 x AC007406/rev ..
Align seg 1/1 to reverse of: AC007406 from: 1 to: 177402
      1 SerAspserlygluThrmetsr 8
      |||||
44426 TCTGATTCTGGGAAACATGTC 44403

seq_name: gb_htg2:AL139019

seq_documentation_block:
LOCUS      AL139019      180140 bp      DNA
DEFINITION Homo sapiens chromosome 1 clone RP5-824018 map p21.3-22.3, ***
ACCESSION      AL139019
VERSION      AL139019.2 GI:7263405
KEYWORDS      HTG; HTGS-PHASE1.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 180140)
AUTHORS      Plumb,B.

```

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (17-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 19, 2000 this sequence version replaced gt:5982957.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known: 800 n's separate
segments. Contig_ID: 00018 length: 10687bp

Contig_ID: 00063 Length: 26542bp
Contig_ID: 00116 Length: 29379bp
Contig_ID: 00246 Length: 25755bp
Contig_ID: 00371 Length: 1183bp
Contig_ID: 00516 Length: 1163bp
Contig_ID: 00617 Length: 5987bp
Contig_ID: 00619 Length: 2768bp
Contig_ID: 00621 Length: 1667bp
Contig_ID: 00689 Length: 3041bp
Contig_ID: 00803 Length: 2384bp
Contig_ID: 00885 Length: 1182bp
Contig_ID: 00998 Length: 32313bp
Contig_ID: 01001 Length: 10594bp
Contig_ID: 01171 Length: 4295bp.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 10687: contig of 10687 bp in length
* 10688 11487: gap of 800 bp
* 11488 38028: contig of 26542 bp in length
* 38030 38829: gap of 800 bp
* 38830 68201: contig of 29372 bp in length
* 68202 69008: gap of 807 bp
* 69009 94763: contig of 25755 bp in length
* 94764 95563: gap of 800 bp
* 95564 96746: contig of 1183 bp in length
* 96747 97546: gap of 800 bp
* 97547 98709: contig of 1163 bp in length
* 98710 99509: gap of 800 bp
* 99510 105496: contig of 5987 bp in length
* 105497 106296: gap of 800 bp
* 106297 109064: contig of 2768 bp in length
* 109065 109864: gap of 800 bp
* 109865 111531: contig of 1667 bp in length
* 111532 112331: gap of 800 bp
* 112332 115372: contig of 3041 bp in length
* 115373 116172: gap of 800 bp
* 116173 118556: contig of 2384 bp in length
* 118557 119356: gap of 800 bp
* 119357 130538: contig of 11182 bp in length
* 130539 131338: gap of 800 bp
* 131339 163651: contig of 32313 bp in length
* 163652 164451: gap of 800 bp
* 164452 175045: contig of 10594 bp in length
* 175046 175845: gap of 800 bp
* 175846 180140: contig of 4295 bp in length.

FEATURES
SOURCE

1. 180140
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p21.3-22.3"
/clone="RP5-824018"
/clone_1lb="RPCT-5"
BASE COUNT 53166 a 30584 c 31027 g 54145 t 11218 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522b-2_COPY_335_428 x AL139019 ..
Align seg 1/1 to: AL139019 from: 1 to: 180140

83 GUGGmMetaGlySThLeuGln 90

170644 GAACAGATGACGAAAGACATTCGAA 170667

seq_name: gb_hgt23:AC012347

seq_documentation_block:

LOCUS AC012347 180435 bp DNA HTG 03-APR-2000
DEFINITION Homo sapiens clone RP11-16G18, WORKING DRAFT SEQUENCE, 4 unordered
pieces.

ACCESSION AC012347
VERSION AC012347.3 GI:7387324

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

2 (bases 1 to 180435)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Batra,N., Beckert,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Daretellano,R., Dewar,R., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hages,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Kiehl,J.,
Lewocky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGuirk,A., McKernan,R., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Strange-Thomson,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tittrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.-J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (25-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 3, 2000 this sequence version replaced gt:6453968.
All repeats were identified using RepeatMasker:
Smt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WTR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3537
Center clone name: 16.G.18
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 169514 bases at least Q40
Consensus quality: 176993 bases at least Q30
Consensus quality: 179409 bases at least Q20
Insert size: 180000; agarose-fp
Insert size: 180135; sum-of-contigs
Quality coverage: 6.5 in Q20 bases; agarose-fp

Quality coverage: 6.5 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 19866: contig of 19866 bp in length

* 19867 19966: gap of 100 bp

* 19967 64174: contig of 44208 bp in length

* 64175 64274: gap of 100 bp

* 64275 122691: contig of 58417 bp in length

* 122692 122792: gap of 100 bp

* 122792 180435: contig of 57644 bp in length.

Location/Qualifiers

1. 180435

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="RPC1-11 Human Male BAC"

1. 19866

/note="assembly_fragment"

19967..64174

/note="assembly_fragment"

clone_end:77

vector_side:left

64275..122691

/note="assembly_fragment"

122792..180435

/note="assembly_fragment"

vector_side:right

clone_end:SP6

BASE COUNT 50782 a 40485 c 39357 g 49511 t 300 others

ORIGIN

alignment_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522b-2_COPY_335_428 x AC012347 ..

Align seg 1/1 to: AC012347 from: 1 to: 180435

46 Valletucysg1yasp1u1a1a 53

|||||

164538 GTTTTACTGTGCGGTGACGAGCA 164561

seq_name: gb_hcg22:AC020581

seq_documentation_block:

LOCUS AC020581 184152 bp DNA HTG 31-MAR-2000

DEFINITION Homo sapiens chromosome 7, *** SEQUENCING IN PROGRESS ***, 47

unorderd pieces.

ACCESSION AC020581

VERSION AC020581.6 GI:7363425

KEYWORDS HTG; HTGS_PHASE1.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.

REFERENCE 1 (bases 1 to 184152)

AUTHORS Olson,M.V.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 184152)

AUTHORS Bubb,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M.

TITLE Direct Submission

JOURNAL

COMMENT

Submitted (05-JAN-2000) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA

On Mar 31, 2000 this sequence version replaced gi:7109511.

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently consists of 47 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 362: contig of 362 bp in length

* 363 725: contig of 363 bp in length

* 726 1433: contig of 708 bp in length

* 1434 2110: contig of 677 bp in length

* 2111 2812: contig of 702 bp in length

* 2813 3199: contig of 387 bp in length

* 3200 3853: contig of 654 bp in length

* 3854 4467: contig of 614 bp in length

* 4468 5120: contig of 633 bp in length

* 5121 5640: contig of 520 bp in length

* 5641 6358: contig of 718 bp in length

* 6359 6784: contig of 426 bp in length

* 6785 7435: contig of 651 bp in length

* 7436 8312: contig of 877 bp in length

* 8313 8874: contig of 562 bp in length

* 8875 9074: contig of 200 bp in length

* 9075 9974: contig of 900 bp in length

* 9975 10598: contig of 624 bp in length

* 10599 11370: contig of 772 bp in length

* 11371 11991: contig of 621 bp in length

* 11992 12857: contig of 866 bp in length

* 12858 13589: contig of 732 bp in length

* 13590 14222: contig of 633 bp in length

* 14223 15031: contig of 809 bp in length

* 15032 15869: contig of 838 bp in length

* 15870 16026: contig of 157 bp in length

gap of unknown length

```

*      16027      16726: contig of 700 bp in length
*      16727      17309: contig of 583 bp in length
*      17310      17791: contig of 482 bp in length
*      17792      18456: contig of 665 bp in length
*      18457      18768: contig of 312 bp in length
*      18769      19142: contig of 374 bp in length
*      19143      20011: contig of 869 bp in length
*      20012      20639: contig of 628 bp in length
*      20640      20995: contig of 356 bp in length
*      20996      21691: contig of 696 bp in length
*      21692      22721: contig of 1030 bp in length
*      22722      23655: contig of 934 bp in length
*      23656      24090: contig of 435 bp in length
*      24091      28092: contig of 4002 bp in length
*      28093      35682: contig of 7590 bp in length
*      35683      47667: contig of 11985 bp in length
*      47668      58578: contig of 10911 bp in length
*      58579      71420: contig of 12842 bp in length
*      71421      93956: contig of 22536 bp in length
*      93957      139756: contig of 45800 bp in length
*      139757      184152: contig of 44396 bp in length.
FEATURES
  source          1..184152
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="7"
BASE COUNT      55675 a 33491 c 35253 g 58554 t 1179 others
ORIGIN
alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
  US-09-332-522b-2_COPY_335_428 x AC020581/rev ..
Align seg 1/1 to reverse of: AC020581 from: 1 to: 184152
      81 HtVAIGUGlMetArGLySThr 88
      |||||
      72275 CATGTTGAGCAGATGAGGAAACT 72252
seq_name: gb_htg10:AC011638
seq_documentation_block:
LOCUS      AC011638      184328 bp      DNA      HTG      01-MAR-2000
DEFINITION Homo sapiens clone Rpl1-13G16, *** SEQUENCING IN PROGRESS ***, 55
ACCESSION      AC011638
VERSION      AC011638.3      GI:7137674
KEYWORDS      HTG; HTGS_PHASE1.

```

```

SOURCE
ORGANISM      human.
REFERENCE
AUTHORS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE        1 (bases 1 to 184328)
JOURNAL      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
REFERENCE    Homo sapiens, clone Rpl1-13G16
AUTHORS      Unpublished
              2 (bases 1 to 184328)
REFERENCE    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
              Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhaltier,B.,
              Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
              Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
              Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
              Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
              Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatsis,A., Klein,J.,
              Lehoczy,J., Lien,C., Locke,K., Macdonald,P., Marquis,N.,
              McEwan,P., McGuirk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
              Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
              Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
              Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
              Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
              Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
              Direct Submission
              Submitted (08-OCT-1999) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
              All repeats were identified using RepeatMasker:
              Smit, A.F.A. & Green, P. (1996-1997)
              http://ftp.genome.washington.edu/XM/RepeatMasker.html
              Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
              Project Information
              Center project name: L3324
              Center clone name: L3_G_16
* NOTE: This is a 'working draft' sequence. It currently
* consists of 55 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1383 1482: contig of 1382 bp in length
1483 2509: contig of 1027 bp in length
2510 2609: gap of 100 bp
2610 3615: contig of 1006 bp in length
3616 3715: gap of 100 bp
3716 5107: contig of 1392 bp in length
5108 5207: gap of 100 bp
5208 6331: contig of 1124 bp in length
6332 6431: gap of 100 bp
6432 7570: contig of 1139 bp in length
7571 7670: gap of 100 bp
7671 8689: contig of 1019 bp in length
8690 8789: gap of 100 bp
8790 9889: contig of 1100 bp in length
9890 9989: gap of 100 bp
9990 11351: contig of 1362 bp in length
11352 11451: gap of 100 bp
11452 13214: contig of 1763 bp in length
13215 13314: gap of 100 bp
13315 14472: contig of 1158 bp in length
14473 14572: gap of 100 bp
14573 16043: contig of 1471 bp in length
16044 16143: gap of 100 bp
16144 17521: contig of 1378 bp in length
17522 17621: gap of 100 bp

```

```

17622 19059: contig of 1438 bp in length
19060 19159: gap of 100 bp
19160 20596: contig of 1437 bp in length
20597 20696: gap of 100 bp
20697 21997: contig of 1301 bp in length
21998 22097: gap of 100 bp
22098 24537: contig of 2440 bp in length
24538 24637: gap of 100 bp
24638 25795: contig of 1158 bp in length
25796 25895: gap of 100 bp
25896 27194: contig of 1299 bp in length
27195 27294: gap of 100 bp
27295 29425: contig of 2131 bp in length
29426 29525: gap of 100 bp
29526 31101: contig of 1576 bp in length
31102 31201: gap of 100 bp
31202 33081: contig of 1880 bp in length
33082 33181: gap of 100 bp
33182 34843: contig of 1662 bp in length
34844 34943: gap of 100 bp
34944 37369: contig of 2426 bp in length
37370 37469: gap of 100 bp
37470 39552: contig of 2083 bp in length
39553 42220: contig of 2568 bp in length
42221 42320: gap of 100 bp
42322 44105: contig of 1785 bp in length
44106 44205: gap of 100 bp
44206 45952: contig of 1747 bp in length
45953 46052: gap of 100 bp
46053 47615: contig of 1563 bp in length
47616 47715: gap of 100 bp
47716 51120: contig of 3405 bp in length
51121 51220: gap of 100 bp
51221 53770: contig of 2550 bp in length
53771 53870: gap of 100 bp
53871 55963: contig of 3093 bp in length
55964 57063: gap of 100 bp
57064 59799: contig of 2736 bp in length
59800 63301: contig of 3402 bp in length
63302 63401: gap of 100 bp
63402 64177: contig of 776 bp in length
64178 64277: gap of 100 bp
64278 66030: contig of 4753 bp in length
66031 69130: gap of 100 bp
69131 73519: contig of 4589 bp in length
73520 73619: gap of 100 bp
73620 76256: contig of 2637 bp in length
76257 76356: gap of 100 bp
76357 80598: contig of 4242 bp in length
80599 80698: gap of 100 bp
80699 84156: contig of 3458 bp in length
84157 84256: gap of 100 bp
84257 89113: contig of 4857 bp in length
89114 89213: gap of 100 bp
89214 93633: contig of 4420 bp in length
93634 93733: gap of 100 bp
93734 98328: contig of 4595 bp in length
98329 98428: gap of 100 bp
98429 102945: contig of 4517 bp in length
102946 103045: gap of 100 bp
103046 108095: contig of 5050 bp in length
108096 108195: gap of 100 bp
108196 114837: contig of 6642 bp in length
114838 114937: gap of 100 bp
114938 121563: contig of 6626 bp in length
121564 121663: gap of 100 bp
121664 128405: contig of 6742 bp in length
128406 128505: gap of 100 bp
128506 133978: contig of 5473 bp in length
133979 134078: gap of 100 bp
134079 138994: contig of 4916 bp in length

```

```

* 138995.139094: gap of 100 bp
* 139095 146606: contig of 7512 bp in length
* 146607 146706: gap of 100 bp
* 146707 154091: contig of 7385 bp in length
* 154092 154191: gap of 100 bp
* 154192 163294: contig of 9103 bp in length
* 163295 163394: gap of 100 bp
* 163395 173882: contig of 10488 bp in length
* 173883 173982: gap of 100 bp
* 173983 184328: contig of 10346 bp in length.
Location/Qualifiers
1. 184328
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-13616"
/clone_id="RPC1-11 Human Male BAC"
1. 1382
/note="assembly-fragment"
1483. 12509
/note="assembly-fragment"
2610. 3615
/note="assembly-fragment"
3716. 5107
/note="assembly-fragment"
5208. 6331
/note="assembly-fragment"
6432. 7570
/note="assembly-fragment"
7671. 8689
/note="assembly-fragment"
8790. 9889
/note="assembly-fragment"
9990. 11351
/note="assembly-fragment"
11452. 13214
/note="assembly-fragment"
13315. 14472
/note="assembly-fragment"
14573. 16043
/note="assembly-fragment"
16144. 17521
/note="assembly-fragment"
17622. 19059
/note="assembly-fragment"
19160. 20596
/note="assembly-fragment"
20697. 21997
/note="assembly-fragment"

```

alignment_scores:

```

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

```

US-09-332-522b-2_COPY_335_428 x AC011638/rev ..

```

```

Align seg 1/1 to reverse of: AC011638 from: 1 to: 184328

```

```

56 SerlySera1aThrLeuArgArg 63
|||||
176984 ACCAAATCCGCAACACTGAGAGG 176961

```

```

seq_name: gb_hlg15:AC016721

```

seq_documentation_block:

```

LOCUS AC016721 185286 bp DNA HTG 13-MAR-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-327J6, WORKING DRAFT SEQUENCE,
19 unordered pieces.
ACCESSION AC016721
VERSION AC016721.4 GI:7230989
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.

```

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 185286)
 AUTHORS Waterston, R.H.
 TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 185286)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (04-DEC-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 COMMENT On Mar 13, 2000 this sequence version replaced gi:7022633.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H_NH0327006
 ----- Summary Statistics -----
 Sequencing vector: M13; 81%
 Sequencing vector: Plasmid; 19%
 Chemistry: Dye-terminator Big Dye; 19% of reads
 Chemistry: Dye-terminator Big Dye; 19% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 173248 bases at least Q40
 Consensus quality: 177028 bases at least Q30
 Consensus quality: 179193 bases at least Q20
 Insert size: 180000; agarose-ff
 Insert size: 183486; sum-of-contigs
 Quality coverage: 4.00 in Q20 bases; agarose-ff
 Quality coverage: 3.92 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
 consists of 19 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1052: contig of 1052 bp in length
 * 1053 1152: gap of unknown length
 * 1153 3846: contig of 2694 bp in length
 * 3847 3946: gap of unknown length
 * 3947 6182: contig of 2236 bp in length
 * 6183 6282: gap of unknown length
 * 6283 11627: contig of 5345 bp in length
 * 11628 11727: gap of unknown length
 * 11728 18380: contig of 6653 bp in length
 * 18381 18480: gap of unknown length
 * 18481 24477: contig of 5997 bp in length
 * 24478 24577: gap of unknown length
 * 24578 31652: contig of 7075 bp in length
 * 31653 31752: gap of unknown length
 * 31753 38667: contig of 6915 bp in length
 * 38668 38767: gap of unknown length
 * 38768 45991: contig of 7224 bp in length
 * 45992 46091: gap of unknown length
 * 46092 53002: contig of 6911 bp in length
 * 53003 53102: gap of unknown length
 * 53103 62477: contig of 9375 bp in length
 * 62478 62577: gap of unknown length
 * 62578 71453: contig of 8876 bp in length
 * 71454 71553: gap of unknown length
 * 71554 78627: contig of 7074 bp in length
 * 78628 78727: gap of unknown length
 * 78728 92545: contig of 13818 bp in length
 * 92546 92645: gap of unknown length
 * 92646 104451: contig of 11806 bp in length
 * 104452 104551: gap of unknown length

104552 121910: contig of 17359 bp in length
 * 121911 122010: gap of unknown length
 * 122011 134398: contig of 12388 bp in length
 * 134399 134498: gap of unknown length
 * 134499 158377: contig of 23879 bp in length
 * 158378 158477: gap of unknown length
 * 158478 185286: contig of 28809 bp in length.
 FEATURES location/Qualifiers
 source 1. 185286
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /clone="RP11-327J6"
 BASE COUNT 59954 a 32166 c 32604 g 58737 t 1825 others
 ORIGIN

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-332-522b-2_COPY_335_428 x AC016721/rev ..
 Align seg 1/1 to reverse of: AC016721 from: 1 to: 185286

54 LysleuserlysearAathleu 61
 |||||
 177557 AAACCTTCAAAAGTGCACCTTGG 177534

seq_name: gp_hcg11:AC008532

seq_documentation_block:
 LOCUS AC008532 188316 bp DNA HTG 18-DEC-1999
 DEFINITION Homo sapiens chromosome 19 clone CIT-HSPC_483111, WORKING DRAFT
 SEQUENCE, 42 unordered pieces.
 AC008532
 VERSION AC008532.3 GI:6601040
 KEYWORDS HTG; PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 188316)
 AUTHORS DOE Joint Genome Institute.
 TITLE DOE Joint Genome Institute.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 188316)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On Dec 18, 1999 this sequence version replaced gi:6165169.

-----Genome Center
 Center: Joint Genome Institute
 Web site: http://www.jgi.doe.gov
 -----Summary Statistics-----
 Estimated insert size: 188316; sum-of-contigs estimation
 Estimated insert size: 156970; agarose-ff estimation
 Quality coverage: 4.44x in Q20 bases; agarose-ff estimation
 Quality coverage: 3.70x in Q20 bases; sum-of-contigs estimation

----- NOTE: This is a 'working draft' sequence. It currently
 consists of 42 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 987: contig of 987 bp in length
 * 987 987: gap of unknown length

```

* 988 2054: contig of 1067 bp in length
* 2055 3518: contig of 1464 bp in length
* 3519 5101: contig of 1583 bp in length
* 5102 6515: contig of 1414 bp in length
* 6516 7963: contig of 1448 bp in length
* 7964 9131: contig of 1168 bp in length
* 9132 10384: contig of 1253 bp in length
* 10385 11547: contig of 1163 bp in length
* 11548 12528: contig of 981 bp in length
* 12529 13627: contig of 1099 bp in length
* 13628 15313: contig of 1686 bp in length
* 15314 17117: contig of 1804 bp in length
* 17118 18733: contig of 1616 bp in length
* 18734 19895: contig of 1162 bp in length
* 19896 20911: contig of 1016 bp in length
* 20912 24704: contig of 3793 bp in length
* 24705 26844: contig of 2140 bp in length
* 26845 29318: contig of 2474 bp in length
* 29319 31060: contig of 1742 bp in length
* 31061 32422: contig of 1362 bp in length
* 32423 34494: contig of 2072 bp in length
* 34495 35618: contig of 1124 bp in length
* 35619 38747: contig of 3129 bp in length
* 38748 40024: contig of 1277 bp in length
* 40025 41529: contig of 1505 bp in length
* 41530 43437: contig of 1908 bp in length
* 43438 46559: contig of 3122 bp in length
* 46560 49676: contig of 3117 bp in length
* 49677 52148: contig of 2472 bp in length
* 52149 56889: contig of 4741 bp in length
* 56890 60983: contig of 4094 bp in length
* 60984 67456: contig of 6473 bp in length
* 67457 73094: contig of 5638 bp in length
* 73095 82758: contig of 9664 bp in length
* 82759 91961: contig of 9203 bp in length
* 91962 102023: contig of 10062 bp in length
* 102024 113956: contig of 11933 bp in length

```

```

* 113957 126563: gap of unknown length
* 126564 140286: gap of unknown length
* 140287 162197: gap of unknown length
* 162198 188316: gap of unknown length
* 188316 198316: Location/Qualifiers
  source
    .188316
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="19"
  /clone="CIT-HSPC-483111"
  /citra="CIT-HSPC-483111"
BASE COUNT 47083 a 48195 c 45826 g 47132 t 80 others
ORIGIN

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x AC008532/rev ..
Align seg 1/1 to reverse of: AC008532 from: 1 to: 188316
56 SerlySera1aThrlleuArgArg 63
|||||
127562 AGCAATCTGCACCTCTCGTAGA 127539

seq_name: gb_hlg15:AC021134

seq_documentation_block:
LOCUS AC021134 189184 bp DNA HTG 13-MAR-2000
DEFINITION Homo sapiens chromosome 4 clone RP11-402D23, WORKING DRAFT
SEQUENCE 30 unordered pieces.
ACCESSION AC021134 GI:7231048
VERSION AC021134.4
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 189184)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189184)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
ON Mar 13, 2000 this sequence version replaced gi:7024000.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0402D23
----- Summary Statistics -----
Sequencing vector: M13; 87%
Sequencing vector: plasmid; 13%
Chemistry: Dye-primer ET; 87% of reads
Chemistry: Dye-terminator Big Dye; 13% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 175660 bases at least Q40
Consensus quality: 179001 bases at least Q30
Consensus quality: 181009 bases at least Q20
Insert size: 199000; agarose-1p

```


Insert size: 186284; sum-of-contigs
Quality coverage: 3.24 in Q20 bases; agarose-fp
Quality coverage: 3.56 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 1223: contig of 1223 bp in length
1224 1323: gap of unknown length
1224 2657: contig of 1334 bp in length
1258 2757: gap of unknown length
1258 3780: contig of 1023 bp in length
1258 3781: gap of unknown length
1258 3880: gap of unknown length
1258 5067: contig of 1187 bp in length
1258 5167: gap of unknown length
1258 5607: contig of 1640 bp in length
1258 6807: gap of unknown length
1258 6907: gap of unknown length
1258 6908: contig of 1743 bp in length
1258 8651: gap of unknown length
1258 8751: contig of 1885 bp in length
1258 10635: gap of unknown length
1258 10735: gap of unknown length
1258 12725: contig of 1990 bp in length
1258 12825: gap of unknown length
1258 12826: contig of 2550 bp in length
1258 15376: gap of unknown length
1258 15476: contig of 2646 bp in length
1258 18121: gap of unknown length
1258 18221: gap of unknown length
1258 20659: contig of 2438 bp in length
1258 20759: gap of unknown length
1258 23142: contig of 2383 bp in length
1258 23242: gap of unknown length
1258 23243: contig of 2493 bp in length
1258 25736: gap of unknown length
1258 25835: gap of unknown length
1258 28779: contig of 2944 bp in length
1258 28879: gap of unknown length
1258 31783: contig of 2904 bp in length
1258 31883: gap of unknown length
1258 34668: contig of 2785 bp in length
1258 34669: gap of unknown length
1258 34769: contig of 4341 bp in length
1258 39210: gap of unknown length
1258 43623: contig of 4414 bp in length
1258 43724: gap of unknown length
1258 43724: contig of 5542 bp in length
1258 49266: gap of unknown length
1258 49366: gap of unknown length
1258 55319: contig of 5954 bp in length
1258 55320: gap of unknown length
1258 55420: contig of 6506 bp in length
1258 61925: gap of unknown length
1258 62026: contig of 8272 bp in length
1258 70297: gap of unknown length
1258 70397: gap of unknown length
1258 79333: contig of 8536 bp in length
1258 79334: gap of unknown length
1258 89628: contig of 10595 bp in length
1258 89728: gap of unknown length
1258 89729: contig of 11459 bp in length
1258 101187: gap of unknown length
1258 101287: gap of unknown length
1258 115262: contig of 13975 bp in length
1258 115263: gap of unknown length
1258 128114: contig of 12752 bp in length
1258 128214: gap of unknown length
1258 128215: contig of 16682 bp in length
1258 144996: gap of unknown length
1258 144997: gap of unknown length
1258 165067: contig of 20071 bp in length
1258 165068: gap of unknown length
1258 189184: contig of 24017 bp in length.
1258 189184: Location/Qualifiers
1258 189184

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-402D23"

BASE COUNT 60365 a 34093 c 33625 g 58157 t 2944 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522b-2_COPY_335_428 x AC021134 ..
Align seg 1/1 to: AC021134 from: 1 to: 189184

84 GlnMetArgLysThrLeuGlnAsn 91
|||||
178317 CAGATGAGGAAACACTCAAAAT 178340

seq_name: gb_in2:AE0015722

seq_documentation_block:
LOCUS AE0015722 193574 bp DNA INV 11-MAR-1999
DEFINITION Drosophila melanogaster Antennapedia complex (ANT-C), section 2 of 2
of the complete sequence.

ACCESSION AE001574
VERSION AE001574.1 GI:4389419

KEYWORDS
SEGMENT
SOURCE
ORGANISM

fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 193574)
2 (bases 1 to 193574)
Martín, C.H., Arcalá, T.T., Bondoc, M.M., Celniker, S.E., Chiang, A.,
Critz, P.A., Davis, C.A., Doyle, C.M., Ericsson, C.L., Farfan, D.E.,
Gunning, K.M., Houston, K.A., Jaklevic, M.A., Kadner, K.E., Kim, K.,
Kim, S.F., Ko, C.L., Lewis, K.D., Li, M., Lindquist, K.J., Lomutan, M.A.,
Lustre, V.M., Machrus, M.U., Mayeda, C.A., Miguel, T.M., Miller, C.A.,
Mok, M.S., Pacled, J.M., Patel, S.G., Pfeiffer, B., Santos, R.F.,
Subramanian, S., Wan, K.H., Whitehead, K.R., Yee, A., Yeh, R.T., Yu, C.
and Palazzolo, M.J.
Direct Submission
Submitted (14-JAN-1999) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, 1 Cyclotron, Berkeley, CA
94720, USA

location/Qualifiers
1. 193574

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
complement(601..684)
/gene="195:84Aa"

gene
trna
complement(601..684)
/product="trna-Lys"
/gene="195:84Aa"
complement(101n(18148..18951,19218..19597,26735..26912,
26978..27078,27464..28250))

gene
complement(18148..28739)
/product="homeodomain protein"
/gene="Deformed"

gene
complement(101n(18637..18951,19218..19597,26735..26912,
26978..27078,27464..28250))
/gene="Deformed"

FEATURES
source

```
/codon_start=1
/product="homodomain protein"
/protein_id="AAd19796.1"
/db_xref="GI:4389426"
/translation="MSSFLMGYPFHAPHVQSPKMGNGLDPKFPPPLADYHHYNGHYS
MNAISTGMGAVGGAGVGVGGAGGAGMTGPHSHMHPADMYQAHANHHNPHSHSHS
HHSLSLPHHNSNSISGHHQASAGYSNTANATPSPHSHPHSHHOSLGYVHNAP
EFTSAGAVSDPTNGYGPANYPNTSGGGSGAVLGAVGASANAGYGGGGY
GANGSVGSHSGHSGPHSQMMDLPQCSTPEPTTALGLDELGKLRLEENPA
GGQDLGMRRLCRDMSGSENDSESDRLMLDSPELGSNDNDLGDSDDEDLMA
ETTDGERILYPMWKILHAGVANGSYQPMERKQRTATRHQILELEEFYNRLL
RRRLRIATLILYSEROIKTFQNRMRKKNKDKNLPNTKARKKTYDANGNPTPYAK
PIKRAASKKQOAOQOQOQOQOOTOOTPYMNECTIRSDLSIEGVSSTGYPYIPAA
PETTSYSGSOOLSNNNNGSGNNNNNNNNNNNNNNNNNNNNNGHNTLHGLHQOQ
SOLMTLQHLIKQDYDTAL"
complement(join(53389..56742,57059..57167))
/mrna
/gene="amalgam"
/product="Immunoglobulin-C2-type-domain protein"
/complement(53389..57167)
/gene="amalgam"
/complement(55734..56735)
/codon_start=1
/gene="amalgam"
/product="Immunoglobulin-C2-type-domain protein"
/protein_id="AAd19797.1"
/db_xref="GI:4389427"
/translation="MARRLILGLIRCLAISDSVLSAPVISOIKDYAVASGVSYER
NCTVEGGLSVSMARKPESDINSVYLSMRNLISLPDRYNTVTEGRTSALYTE
RQNIIEVSDMGPEECVLSATEKVKRLSLQIKTPPYAENTPKSTLVEQNGELT
CHANGPKPTISMAREHNAVMPAGHLAEPILIRSVHRMDRGYYCIAQNGEGPD
KRLIRVEFERPQIAVORPKIAMQVSHSALECSYGVGAPATVYMKNGVPIOSSHH
EVANTSSSGSTTSYLRIDSGEDEGDYCNANTNLGHADARHLHFCVIVPLSL"
join(61135..61468,62028..62103,62144..63102,63615..64655)
/mrna
/gene="bicoloid"
/product="DNA-binding protein, transcription factor"
/gene="bicoloid"
join(61304..61468,62028..62103,62144..63102,63615..63899)
/codon_start=1
/gene="bicoloid"
/product="DNA-binding protein, transcription factor"
/protein_id="AAd19798.1"
/db_xref="GI:4389428"
/translation="MAOPRPDQNFYHNRLPHTHTPHRPHSHRPHSHRHHOQLO
LPORNPRLDLFDERTGAINYNYIRPYIPNOMPKRDVPSSELPDSLYMRPKRRT
PTTSQIALEQHFQGLRILTAFLADLSAKLALGTAQVIMFKNRRRHKLIQSDQHK
DQSECPMLSPGKQSDGDPPLQTLISLGGAIPNALTPSPSTPTAHMTBYSSEF
NAYYNGGHNQAQNRHMQYPSGGGPGSTYNGGQFOQOOVHNHQOOLHHQG
NHYPHMQOQOQOAOQOQYHHFDFQOKASACRYLKDEPADYFNNSYSYMRSGSG
ATASASAVARGAASRSEYERPLTPKNDSPSLCGIGGPCAIANGTEADDDMDG
TSKTTQLLEPLKLGIDKSDGSSSDMSTGIRALAGTCNKGAFKFKGPPSPQPO
PFLGGMVAGMSNDQCTMDYTMQATYNHRNNAAGNSOPAYCFN"
<65610..66284
/gene="ORF1"
<65616..66284
/gene="ORF1"
/codon_start=1
/product="unknown"
/protein_id="AAd19799.1"
/db_xref="GI:4389429"
/translation="RYFOLYLTLAATFPGYVPMPLFORIFSEFEGENPIAVYIK
FANSLYVGLRMGRPMDRPDTSDSHSRRLRLHILLGNQSKLSKSSSKPPGASLT
AALSLHENCRCRRAKSRDNSFGFTFSATGGSFTRCKLNNRLGSEITLISPSIWR
DSGSGCPGTLELSTQRLLAGFRSKQASRLPLANGRESSDLASLGASRAVORRDL
RLAA"
join(66376..66513,66578..67724)
/mrna
/gene="zerknüllt 1"
/product="DNA-binding protein, transcription factor"
/gene="zerknüllt 1"
join(66427..66513,66578..67552)
/gene="zerknüllt 1"
/codon_start=1
```

```
/product="DNA-binding protein, transcription factor"
/protein_id="AAd19800.1"
/db_xref="GI:4389430"
/translation="MSSVMHYYPVHQARVGSADPSEVKYSDLYGHHHDVNPICLP
PNYNQMNPNPTLNDHCSPQHVQOHSVSDENLPQPNHDSQRYVLRKSRATFVOL
VLENEFNSNMYLYTRRLIEAQRSLCEQVRIKWFONRRMKFKKIDGHEPKSNK
LAQPAQESADHGIYKRLMSYSQDREETAERKPMNAVAVPNVKNPYOASOKKTE
ASTNGMSANLDSITLHLMOTTAAPVSRATSTGTSNASSSGHSTYNDLV
LOSTIKODLEAQAQMSKRSAPILATQSHPSQSQVPTSVHARSNMLSWGEPAKS
KRLSYHNHNPCTSYNYPN"
join(76435..76560,76628..77443)
/mrna
/gene="zerknüllt 2"
/product="DNA-binding protein, transcription factor"
/gene="zerknüllt 2"
join(76435..76560,76628..77326)
/gene="zerknüllt 2"
/codon_start=1
/product="DNA-binding protein, transcription factor"
/protein_id="AAd19801.1"
/db_xref="GI:4389431"
/translation="MFAIQSENYFYDNTSVSDLMYPCVELNEVAPATRTSRSEK
RSRTAFSSLOILELEERFHLNKLARTIRISORLALTEQVAKIWFONRRMKLKST
NRKGAIGALTSIPSSQSSDLOKDDQIVRLIRYANTNETAPLROVDVDEGQ
ITPPYQSDYLTHERSPERPMALPOLPFNEFDMWMSWMLGELPTPIAENYIENHQD
PHIQFCNDSNSSSSSDITDVIDYDFONTLNF"
join(78512..79407,80401..81025,93343..93357,
108624..108781,108956..109070,109613..109744,
109801..109973,110201..110958,111919..112778)
/gene="proboscipedia"
/product="homodomain protein"
/gene="proboscipedia"
join(78512..112778)
/gene="proboscipedia"
join(80476..81025,93343..93357,108624..108781,
108956..109070,109613..109744,109801..109973,
110201..110958,111919..112366)
/gene="proboscipedia"
/codon_start=1
/product="homodomain protein"
/protein_id="AAd19802.1"
/db_xref="GI:4389432"
/translation="MQEYCSLDITTSWGTQIKSSPPLNPLOVOTGOTSLPYGGGGAG
VVGQYGVGVGVGQPGTGOQGVGPPVPSYILGNKMTPNCKDRKSAPTAYMTASEGFIN
SQPSAEFLNHSRSPERKIGTPVSGAIGYGVNVNVAVGVGIPGVVQTPDGGG
SVPEYPMKERRKTSKSSNNNQDGNSTIEVVPENGRLRLRTAYTNOLLEKEFH
FNKYICRARRIEIAASDLITEROVYVQNRNKHKROTLSKTDDEKDSLKGDDQ
SDSNSKSCQGCPLPSDDIPDSTNSGRHNNTPSATNNPSAGNLTPSSLETGI
SNLMGSTYSASNYISADSSVASVSLDEDIIESSPIKVKKDDGVYIKKEAVSTSS
KASPGYNSNPSIYSPFRDSDASVGNAPTSKAVKRPQSANALITPTPLSDNS
GNGSGGAGATFPGYTPSPKQOQVQOQQLHPQOQLPQOQPODYTGKDIERFAS
HNPNKQOQALHGEYLSRPSANEHQSDQOQNDHYTYUNODTNGPXYLNHQOQH
HHAQHQQOQHONNVADPEGVNPSFNNGAAYDNNFQOQALHGHQVVFQOQ
PHQPAIINHONHNLHNGEYLSALGLOENEGVNFQOQALGCGGYTAGOOPPIAT
HGHGHNPHVOVPAQAHAIHANSALIPGQYGVGPPSPSHHGALNGRAVYQCAF
GNGSTAGTAIISGLEMSNSDPRFLSNLANDRAPETQLS"
complement(join(115755..116427))
/gene="Ccp84Aa"
/gene="Ccp84Aa"
complement(join(115755..116360,116416..116427))
/gene="Ccp84Ab"
/codon_start=1
/product="cuticle protein"
/protein_id="AAd19803.1"
/db_xref="GI:4389433"
/translation="MAEFVVALAFVAVASGVAPIAPQYVHAAPAVATYHAAPVAV
AQVYVKAEEYDNPQYRFSGYVDKTLGKNKGQVDEGDDVYGEYSLIDAGYKR
IVQYADIDINEFNVNREPLVKAVAVAPVKTVAAPVQYAAAPVAAIPVAVYKTV
APVAAVAPVAYKTVAPVAAVAAAPVATYTAAPVHYAAPVAYHP"
117755..118496
/gene="Ccp84Ab"
join(117755..117766,117843..118496)
/gene="Ccp84Ab"
/codon_start=1
/product="cuticle protein"
```

/protein_id="A019804.1"
/db_xref="GI:4389434"
/translation="MAKFFVFALAFYAVASAGTAPAPDQVTHAPAVATYAHAPVAV
ADKVVVKAEEYDHPQYRFSYSGVDKLTGDNKQVEERGDVREYSIIDADGYSKR
TVQYADPINGENAVNREPLVAVAPVYKAVAPVAQVAPAVAHAAAPVYTV
APVAVHAAPAVVKTAVPAVHAHYAPAPVATYAPATHAAAPAHYAPATATTSYAPAPA

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

_US-09-332-522b-2_COPY_335_428 x AE0015722/rev ..

Align seg 1/1 to reverse of: AE0015722 from: 1 to: 193574

54 LysleuSerLysSerAlaThrIleu 61

178952 AAACGTCCAAAGTCGCGACTCTT 178929

seq_name: gb_hlg28:AC009329

seq_documentation_block:

LOCUS AC009329 197530 bp DNA HTG 19-APR-2000
DEFINITION Homo sapiens chromosome 7, *** SEQUENCING IN PROGRESS ***, 42
unoriented pieces.

ACCESSION AC009329
VERSION AC009329.18 GI:7596786
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 197530)
Olson,M.V.

REFERENCE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 197530)
AUTHORS Bubb,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M.
TITLE Direct Submission

JOURNAL Submitted (16-AUG-1999) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
On Apr 19, 2000 this sequence version replaced gi:7363426.

COMMENT

Sequence Quality Assessment:

This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.

NOTE: This is a 'working draft' sequence. It currently
consists of 42 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 859: contig of 859 bp in length
* 860 1032: gap of unknown length
* 1033 1866: gap of unknown length
* 1867 2349: gap of unknown length
* 2350 3218: gap of unknown length
* 3219 3864: gap of unknown length
* 3864: contig of 646 bp in length

FEATURES

3865 4648: gap of unknown length
* 4649 5749: gap of unknown length
* 5750 6855: gap of unknown length
* 6856 7774: gap of unknown length
* 7775 8689: gap of unknown length
* 8690 9629: gap of unknown length
* 9630 10453: gap of unknown length
* 10454 11120: gap of unknown length
* 11121 12062: gap of unknown length
* 12063 12713: gap of unknown length
* 12714 13563: gap of unknown length
* 13564 15150: gap of unknown length
* 15151 15261: gap of unknown length
* 15262 15920: gap of unknown length
* 15921 16662: gap of unknown length
* 16663 17412: gap of unknown length
* 17413 17549: gap of unknown length
* 17550 18120: gap of unknown length
* 18121 19226: gap of unknown length
* 19227 20149: gap of unknown length
* 20150 21302: gap of unknown length
* 21303 22648: gap of unknown length
* 22649 23667: gap of unknown length
* 23668 24973: gap of unknown length
* 24974 25681: gap of unknown length
* 25682 25917: gap of unknown length
* 25918 27331: gap of unknown length
* 27332 28223: gap of unknown length
* 28224 29178: gap of unknown length
* 29179 30459: gap of unknown length
* 30460 31234: gap of unknown length
* 31235 32135: gap of unknown length
* 32136 33660: gap of unknown length
* 33661 35293: gap of unknown length
* 35294 36776: gap of unknown length
* 36777 197530: gap of unknown length
Location/Qualifiers

```

source
1..197530
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"

BASE COUNT 60806 a 37739 c 37816 g 61091 t 78 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522b-2_COPY_335_428 x AC008591 ..
Align seg 1/1 to: AC008591 from: 1 to: 197530

81 Hisva1gluglunmetarglystr 88
|||||
183867 CATGTGCGAGATGAGGAAACT 183890

seq_name: gb_hlg12:AC008591

seq documentation block:
LOCUS AC008591 199199 bp DNA HTG 18-DEC-1999
DEFINITION Homo sapiens chromosome 5 clone CIT-HSPC_575N7, WORKING DRAFT
SEQUENCE 32 unordered pieces.
ACCESSION AC008591
VERSION AC008591.3 GI:6601023
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 199199)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 199199)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Dec 18, 1999 this sequence version replaced gi:6165161.
-----Genome Center
Center: Joint Genome Institute
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Estimated insert size: 189199; sum-of-contigs estimation
Estimated insert size: 188000; pulse field gel estimation
Quality coverage: 4.34x in Q20 bases; pulse field gel estimation
Quality coverage: 4.10x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved
*
* 1 1006: contig of 1006 bp in length
* gap of unknown length
* 1007 2342: contig of 1336 bp in length
* gap of unknown length
* 2343 3720: contig of 1378 bp in length
* gap of unknown length
* 3721 4795: contig of 1075 bp in length
* gap of unknown length
* 4796 5850: contig of 1055 bp in length
* gap of unknown length
* 5851 7393: contig of 1543 bp in length

```

```

*
* 7394 9467: gap of unknown length
* contig of 2074 bp in length
* 9468 11357: gap of unknown length
* contig of 1890 bp in length
* 11358 12771: gap of unknown length
* contig of 1414 bp in length
* 12772 14334: gap of unknown length
* contig of 1563 bp in length
* 14335 15849: gap of unknown length
* contig of 1515 bp in length
* 15850 18235: gap of unknown length
* contig of 2386 bp in length
* 18236 21621: gap of unknown length
* contig of 3386 bp in length
* 21622 24606: gap of unknown length
* contig of 2985 bp in length
* 24607 27319: gap of unknown length
* contig of 2713 bp in length
* 27320 30996: gap of unknown length
* contig of 3677 bp in length
* 30997 34133: gap of unknown length
* contig of 3137 bp in length
* 34134 39402: gap of unknown length
* contig of 5269 bp in length
* 39403 44715: gap of unknown length
* contig of 5313 bp in length
* 44716 50100: gap of unknown length
* contig of 5385 bp in length
* 50101 55246: gap of unknown length
* contig of 5146 bp in length
* 55247 60890: gap of unknown length
* contig of 5644 bp in length
* 60891 66816: gap of unknown length
* contig of 7726 bp in length
* 66817 74648: gap of unknown length
* contig of 6032 bp in length
* 74649 82809: gap of unknown length
* contig of 8161 bp in length
* 82810 95991: gap of unknown length
* contig of 13182 bp in length
* 95992 106311: gap of unknown length
* contig of 10320 bp in length
* 106312 118931: gap of unknown length
* contig of 12620 bp in length
* 118932 134217: gap of unknown length
* contig of 15286 bp in length
* 134218 148153: gap of unknown length
* contig of 13936 bp in length
* 148154 169798: gap of unknown length
* contig of 21645 bp in length
* 169799 199199: gap of unknown length
* contig of 29401 bp in length.
* Location/Qualifiers
* 1..199199
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="5"
* /clone="CIT-HSPC_575N7"

BASE COUNT 62721 a 36702 c 37137 g 62238 t 391 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-332-522b-2_COPY_335_428 x AC008591/rev ..
Align seg 1/1 to reverse of: AC008591 from: 1 to: 199199
54 Lysleuserlyseralathreu 61

```

114531 AACCTTCTACAGATGCCACCTG 114508

seq_name: gb_hgt18:AC026782

seq_documentation_block:

LOCUS AC026782 213902 bp DNA HTG 24-MAR-2000

DEFINITION Homo sapiens chromosome 5 clone CTD-2015A6, WORKING DRAFT SEQUENCE.

28 unordered pieces.

ACCESSION AC026782

VERSION AC026782.1 GI:7321427

KEYWORDS HTG: HTGS_PHASEL; HTGS_DRAFT.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 213902)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 5

Unpublished

2 (bases 1 to 213902)

DOE Joint Genome Institute.

Direct Submission

Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: http://www.jgi.doe.gov

-----Summary Statistics

Consensus quality: 180358 bases at least Q40

Consensus quality: 200699 bases at least Q20

Consensus quality: 206898 bases at least Q20

Estimated insert size: 213902; sum-of-ctgfs estimation

Estimated insert size: 218000; pulse field gel estimation

Quality coverage: 4.09x in Q20 bases; pulse field gel estimation

Quality coverage: 4.17x in Q20 bases; sum-of-ctgfs estimation

NOTE: This is a 'working draft' sequence. It currently

consists of 28 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 1054: contig of 1054 bp in length
gap of unknown length
1055 2351: contig of 1297 bp in length
gap of unknown length
2352 3410: contig of 1059 bp in length
gap of unknown length
3411 5079: contig of 1669 bp in length
gap of unknown length
5080 6113: contig of 1034 bp in length
gap of unknown length
6114 7353: contig of 1240 bp in length
gap of unknown length
7354 9273: contig of 1920 bp in length
gap of unknown length
9274 10851: contig of 1578 bp in length
gap of unknown length
10852 12731: contig of 1880 bp in length
gap of unknown length
12732 15741: contig of 3010 bp in length
gap of unknown length
15742 19008: contig of 3267 bp in length
gap of unknown length
19009 22025: contig of 3017 bp in length
gap of unknown length
22026 23701: contig of 1676 bp in length
gap of unknown length
23702 25978: contig of 2277 bp in length

* 25979 28200: gap of unknown length
contig of 2222 bp in length
* 28201 31465: gap of unknown length
contig of 3265 bp in length
* 31466 37915: gap of unknown length
contig of 6450 bp in length
* 37916 45180: gap of unknown length
contig of 7265 bp in length
* 45181 57905: gap of unknown length
contig of 12725 bp in length
* 57906 66505: gap of unknown length
contig of 8600 bp in length
* 66506 76981: gap of unknown length
contig of 10476 bp in length
* 76982 88626: gap of unknown length
contig of 11645 bp in length
* 88627 94572: gap of unknown length
contig of 5946 bp in length
* 94573 110548: gap of unknown length
contig of 15976 bp in length
* 110549 129319: gap of unknown length
contig of 18771 bp in length
* 129320 157096: gap of unknown length
contig of 27777 bp in length
* 157097 183356: gap of unknown length
contig of 26260 bp in length
* 183357 213902: gap of unknown length
contig of 30546 bp in length.
Location/Qualifiers
1. 213902
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2015A6"

BASE COUNT 65678 a 42019 c 41232 g 64881 t 92 others
ORIGIN

Alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment_block:

US-09-332-522b-2_COPY_335_428 x AC026782/rev ..

Align seg 1/1 to reverse of: AC026782 from: 1 to: 213902

58 SerialThrlenuArgAlaIle 65

82618 TCAGCCACATGTGAGGAGGCAATC 82595

seq_name: gb_hgt2:AL139331

seq_documentation_block:

LOCUS AL139331 227106 bp DNA HTG 16-FEB-2000

DEFINITION Homo sapiens chromosome 6 clone RP11-338M4, *** SEQUENCING IN

PROGRESS ***, 37 unordered pieces.

ACCESSION AL139331

VERSION AL139331.2 GI:6996191

KEYWORDS HTG: HTGS_PHASEL.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 227106)

Slms,S.

Direct Submission

Submitted (16-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humgeny@sanger.ac.uk

Requests: clonerequests@sanger.ac.uk

On Feb 17, 2000 this sequence version replaced g1:6996067.

IMPORTANT: This sequence is unfinished and does not necessarily

COMMENT

represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Contig_ID: 00002 Length: 7355bp

```
Contig_ID: 00036 Length: 1483bp
Contig_ID: 00037 Length: 4262bp
Contig_ID: 00047 Length: 2244bp
Contig_ID: 00063 Length: 4296bp
Contig_ID: 00094 Length: 3655bp
Contig_ID: 00165 Length: 1398bp
Contig_ID: 00182 Length: 3287bp
Contig_ID: 00204 Length: 1054bp
Contig_ID: 00212 Length: 1315bp
Contig_ID: 00398 Length: 1025bp
Contig_ID: 00407 Length: 2754bp
Contig_ID: 00411 Length: 6502bp
Contig_ID: 00422 Length: 2122bp
Contig_ID: 00554 Length: 1325bp
Contig_ID: 00572 Length: 1056bp
Contig_ID: 00605 Length: 1371bp
Contig_ID: 00626 Length: 1057bp
Contig_ID: 00637 Length: 9972bp
Contig_ID: 00833 Length: 8597bp
Contig_ID: 00851 Length: 1112bp
Contig_ID: 00924 Length: 1050bp
Contig_ID: 00942 Length: 15290bp
Contig_ID: 00988 Length: 9654bp
Contig_ID: 01027 Length: 15183bp
Contig_ID: 01033 Length: 1030bp
Contig_ID: 01042 Length: 2297bp
Contig_ID: 01076 Length: 10118bp
Contig_ID: 01112 Length: 2190bp
Contig_ID: 01124 Length: 2969bp
Contig_ID: 01129 Length: 4713bp
Contig_ID: 01270 Length: 3791bp
Contig_ID: 01313 Length: 8268bp
Contig_ID: 01348 Length: 2438bp
Contig_ID: 01507 Length: 2168bp
Contig_ID: 01514 Length: 10340bp
Contig_ID: 01564 Length: 9349bp.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
7356 8155: gap of 800 bp in length
8156 9638: contig of 1483 bp in length
9639 10438: gap of 800 bp
10439 14700: contig of 4262 bp in length
14701 15500: gap of 800 bp
15501 37944: contig of 22444 bp in length
37945 38744: gap of 800 bp
38745 43040: contig of 4296 bp in length
43041 43840: gap of 800 bp
43841 47495: contig of 3655 bp in length
47496 48295: gap of 800 bp
48296 49693: contig of 1398 bp in length
49694 50493: gap of 800 bp
50494 53780: contig of 3287 bp in length
53781 54580: gap of 800 bp
54581 55634: contig of 1054 bp in length
55635 56434: gap of 800 bp
56435 57749: contig of 1315 bp in length
57750 58549: gap of 800 bp
58550 59574: contig of 1025 bp in length
59575 60374: gap of 800 bp
60375 63128: contig of 2754 bp in length
```

```
* 63129 63928: gap of 800 bp
* 63929 70430: contig of 6502 bp in length
* 70431 71230: gap of 800 bp
* 71231 73352: contig of 2122 bp in length
* 73353 74152: gap of 800 bp
* 74153 75477: contig of 1325 bp in length
* 75478 76277: gap of 800 bp
* 76278 77333: contig of 1056 bp in length
* 77334 78133: gap of 800 bp
* 78134 79504: contig of 1371 bp in length
* 79505 80304: gap of 800 bp
* 80305 81361: contig of 1057 bp in length
* 81362 82161: gap of 800 bp
* 82162 92133: contig of 9972 bp in length
* 92134 92933: gap of 800 bp
* 92934 101530: contig of 8597 bp in length
* 101531 102330: gap of 800 bp
* 102331 113454: contig of 11124 bp in length
* 113455 114254: gap of 800 bp
* 114255 115304: contig of 1050 bp in length
* 115305 116104: gap of 800 bp
* 116105 131394: contig of 15290 bp in length
* 131395 132194: gap of 800 bp
* 132195 141848: contig of 9654 bp in length
* 141849 142648: gap of 800 bp
* 142649 157837: contig of 15189 bp in length
* 157838 158637: gap of 800 bp
* 158638 159667: contig of 1030 bp in length
* 159668 160467: gap of 800 bp
* 160468 162764: contig of 2297 bp in length
* 162765 163564: gap of 800 bp
* 163565 173682: contig of 10118 bp in length
* 173683 174482: gap of 800 bp
* 174483 176672: contig of 2190 bp in length
* 176673 177472: gap of 800 bp
* 177473 180441: contig of 2969 bp in length
* 180442 181241: gap of 800 bp
* 181242 185954: contig of 4713 bp in length
* 185955 186734: gap of 800 bp
* 186735 190545: contig of 3791 bp in length
* 190546 191345: gap of 800 bp
* 191346 199613: contig of 8268 bp in length
* 199614 200413: gap of 800 bp
* 200414 202851: contig of 2438 bp in length
* 202852 203651: gap of 800 bp
* 203652 205817: contig of 2166 bp in length
* 205818 206617: gap of 800 bp
* 206618 216957: contig of 10340 bp in length
* 216958 217757: gap of 800 bp
* 217758 227106: contig of 9349 bp in length.
* Location/Qualifiers
```

```
source
1..227106
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-328M4"
/clone_lib="RP11-11.2"
BASE COUNT 48239 a 50617 c 50963 g 48435 t 28852 others
```

```
ORIGIN
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-332-522b-2_COPY_335_428 x AL139331 ..
```

```
Align seg 1/1 to: AL139331 from: 1 to: 227106
```

```
25 ALH1SASnleu1leg1uylslys 32
|||||
```

```
45899 GCACACACCTCATGTGAGAAAAAG 45922
seq_name: gb_in2:AE003674
seq_documentation_block:
LOCUS      AE003674      298028 bp      DNA      INV      22-MAR-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386040 section 4
            of 5 complete sequence.
ACCESSION  AE003674 AE002659
VERSION     AE003674.1 GI:7298877
KEYWORDS    HTG.
SOURCE      fruit fly.
            Drosophila melanogaster
ORGANISM    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 298028)
            Adams,M.D., Celiker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
AUTHORS     Amaratides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazer,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews,G.L., Baskin,C., Baldwin,D., Baller,R.M., Basu,A.,
            Barendse,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borokova,D.,
            Botchan,M.R., Bouck,J.J., Brokstein,P., Brotlier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablo,B., Delcher,A., Deng,Z., Mays,A.D., Dev,I., Dietz,S.M.,
            Doonan,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferrara,C., Fertiler,A.S.,
            Fleischmann,W., Fosler,K., Gabori,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Idegawa,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kienison,J.A., Ketchum,K.A., Kimmel,B.E., Kodera,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lascko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazolo,M., Peltman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spredling,A.C.,
            Stopleton,M., Strong,R., Sun,E., Svirskas,R., Tector,G., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,Q., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Glibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
            The genome sequence of Drosophila melanogaster
            Science 287 (5461), 2185-2195 (2000)
TITLE       2 (bases 1 to 298028)
JOURNAL     20196006
MEDLINE     Science 287 (5461), 2185-2195 (2000)
REFERENCE   1 (bases 1 to 298028)
AUTHORS     Adams,M.D., Celiker,S.E., Glibbs,R.A., Rubin,G.M. and Venter,J.C.
JOURNAL     Direct Submission
            Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
FEATURES    Location/Qualifiers
            source          1..298028
                                /organism="Drosophila melanogaster"
                                /db_xref="taxon:7227"
                                /chromosome="3R"
                                complement(join(13398..14406,14723..>14823))
                                /gene="Ama"
                                /product="CT7244"
                                /db_xref="FlyBASE:FBan0002198"
                                /db_xref="FlyBASE:FBgn0000071"
                                /map="84A5-84A5"
gene        complement(<13398..>14823)
            /gene="Ama"
            /note="CG2198"
            /map="84A5-84A5"
            /db_xref="FlyBASE:FBan0002198"
            /db_xref="FlyBASE:FBgn0000071"
            complement(13398..14399)
            /gene="Ama"
            /note="Ama gene product"
            /codon_start=1
            /db_xref="FlyBASE:FBan0002198"
            /db_xref="FlyBASE:FBgn0000071"
            /protein_id="AAFS4086.1"
            /db_xref="GI:7298878"
            /translation="MARLLRLGLIFCLAISLDSVLSAPVISOIKDVASVGDVSEF
            NCVEEGLQSVMAKRPSSDINSVLSNRNLSLPDQRYNTVTEGPKTGSAYTF
            RQINEVSDMGPEYCEVYATKVTKLSLQIKTPVIAENTPVTKEGQNELT
            CHANGFEKPTISMAREHNVMPAGHLLAEPTLRIRSVHMDGGYCCIANONGEOPD
            KRLRVFERPRQIANGRKIAQVNSASALEGSCVQTPAPTYMHKNGPLSSRRH
            EVANTASSGTTTSLRIDSVGEDEFQDTCNATNKIGHADARLHLFQTVIPVPSLS"
            join(18797..19132,19808..20766,21279..22380)
            /gene="bcd"
            /note="Transcript structure is suspect."
            /product="CT1104"
            /db_xref="FlyBASE:FBan0001034"
            /db_xref="FlyBASE:FBgn0000166"
            <18797..>22380
            /gene="bcd"
            /note="CG1034"
            /map="84A5-84A5"
            /db_xref="FlyBASE:FBan0001034"
            /db_xref="FlyBASE:FBgn0000166"
            join(18968..19132,19808..19846)
            /gene="bcd"
            /note="bcd gene product: Transcript structure is suspect"
            /codon_start=1
            /db_xref="FlyBASE:FBan0001034"
            /db_xref="FlyBASE:FBgn0000166"
            /protein_id="AAFS4085.1"
            /db_xref="GI:7298879"
            /translation="MAQPPDQNYHHNPLHTHTNPHSHPHSHPHNHNQPDQ
            LPFRFRPDLMSPPQASCPFLM"
            23280..>23948
            /gene="Bg:DS00276.8"
            /product="CT2039"
            /db_xref="FlyBASE:FBan0001162"
            /db_xref="FlyBASE:FBgn0026564"
            <23280..>23948
            /gene="Bg:DS00276.8"
            /note="CG1162"
            /db_xref="FlyBASE:FBan0001162"
            /db_xref="FlyBASE:FBgn0026564"
            23337..23948
            /gene="Bg:DS00276.8 gene product"
            /note="Bg:DS00276.8 gene product"
            /codon_start=1
            /db_xref="FlyBASE:FBan0001162"
            /db_xref="FlyBASE:FBgn0026564"
            /protein_id="AAFS4086.1"
            /db_xref="GI:7298880"
            /translation="WTFPLQRTFSFSEGEENPIAPVIRKANNSLVGLRGMGRMR
            PETHSHPRRLLEHILGKNOSKLSKSNKPRGASFLTALLSLHYDNGGRATKSR
            DNSRGFTSATGSGFNRCYKLDNNRKGSPITISPSYWDGSGCGPTELSLTOROI
            LAGFRSKOASRLPIANGRESLDSLGAASAVORRRDLRLAA"
            join(<24091..24177,24242..>25216)
            /gene="zen"
            /product="CT1179"
            /db_xref="FlyBASE:FBan0001046"
            /db_xref="FlyBASE:FBgn0004053"
            <24091..>25216
            /gene="zen"
            /note="CG1046"
            /map="84A5-84A5"
```

```
/db_xref="FLYBASE:FBan0001046"
/db_xref="FLYBASE:FBgn0004053"
join(24091..24177,24242..25216)
/gene="zen"
/note="zen gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0001046"
/db_xref="FLYBASE:FBgn0004053"
/protein_id="AA54087.1"
/db_xref="GI:7298881"
/translation="MSSVMHYYPVHQAKGYSYADPSEVKYSDDLIGHHHNVPICLP
PYNQNMNPTLLNDHCSPQHVCHSSDENLPSQNHDSQVKKRKRSRTAFSTVL
VELENEKSNMTLIRTRRIEIQRLSCEROYKIMQNRMRKRIQGHREPKSAK
LAQPOAESAHGIVIRKMSYODPREGTAAAEKPMMAVAPNRPDPQAOQKMTK
ASTNNKSCSADLSELEHLAOTTAAPQVSTATSGTSTNSHSSSHYSTNDLV
LOSIRKDLAAQAWSKSAPLTAQVSHSPSSQSVTVHAAPSMNLWSPEPAKS
RLISVNHMPCVTSYNYPN"
join(<34165..34224,34292..>34990)
/gene="zen2"
/product="Ct1183"
/db_xref="FLYBASE:FBan0001048"
/db_xref="FLYBASE:FBgn0004054"
<34165..>34990
/gene="zen2"
/note="CG1048"
/map="84A5-84A5"
/db_xref="FLYBASE:FBan0001048"
/db_xref="FLYBASE:FBgn0004054"
join(34165..34224,34292..34990)
/gene="zen2"
/note="zen2 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0001048"
/db_xref="FLYBASE:FBgn0004054"
/protein_id="AA54088.1"
/db_xref="GI:7298882"
/translation="MFAIOSNFVNDVSVDIMYPCVELNVEAPATATTSSEKX
RSRTASSQLIELEPREFLNKIATRIEISQRLALEROYKIFQNRRLKIKST
NKGAIQALITISPLSSOSSEDQKDDQVERLLRYANVETAPLRQVDHGLEGO
ITPPOSYDLHEFSEPMALPOLPNEFDANWASSMLPTPIAENVIEHNTODQ
PMIONFCWDSNSSASSDILVDYDFIONLNF"
join(36176..37071,38065..38689,66288..66445,66620..66734,
67277..67408,67465..67637,67865..68622,69563..70442)
/gene="pb"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CT39787"
/db_xref="FLYBASE:FBan0017880"
/db_xref="FLYBASE:FBgn0003040"
<36176..>70442
/gene="pb"
/note="CG17880"
/map="84A5-84A5"
/db_xref="FLYBASE:FBan0017880"
/db_xref="FLYBASE:FBgn0003040"
join(38140..38689,66288..66445,66620..66734,67277..67408,
67465..67637,67865..68622,69563..70030)
/gene="pb"
/note="pb gene product; Nucleotide sequence of the Celera

alignment_scores:
  Quality:      8 00      Length:      8
  Ratio:        1.000      Gaps:        0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-332-522B-2_COPY_335_428 x AE003674/rev
..
Align seg 1/1 to reverse of: AE003674 from: 1 to: 298028

54 LysleuSerIysSerAlaThrIleu 61
|||||
136616 AAACGTGTCAGAGTCGCGACTCTT 136593
```

```
seq_name: gb_htg19:AC025242
seq_documentation_block:
LOCUS      AC025242      318640 bp      DNA
DEFINITION Homo sapiens clone Rpl1-500C12, *** SEQUENCING IN PROGRESS ***, 108
unnumbered pieces.
ACCESSION  AC025242
VERSION    AC025242.2   GI:7321997
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Euteria; Primates; Carnivora; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 318640)
            Watson,R.H.
            The sequence of Homo sapiens clone
            Unpublished
            2 (bases 1 to 318640)
            Watson,R.H.
            Direct Submission
            Submitted (07-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On Mar 24, 2000 this sequence version replaced gi:7188920.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Project name: H_NH0500C12
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 108 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
1404      1403: contig of 1403 bp in length
1523      1522: gap of unknown length
3083      3082: contig of 1561 bp in length
3084      3202: gap of unknown length
3203      5121: contig of 1919 bp in length
5122      5240: gap of unknown length
5241      6481: contig of 1241 bp in length
6482      6600: gap of unknown length
6601      8027: contig of 1427 bp in length
8028      8146: gap of unknown length
8147      9323: contig of 1177 bp in length
9324      9442: gap of unknown length
9443      10788: contig of 1346 bp in length
10789     10907: gap of unknown length
10908     12100: contig of 1193 bp in length
12101     12219: gap of unknown length
12220     13513: contig of 1294 bp in length
13514     13632: gap of unknown length
13633     15286: contig of 1654 bp in length
15287     15405: gap of unknown length
15406     17104: contig of 1699 bp in length
17105     17223: gap of unknown length
17224     18736: contig of 1513 bp in length
18737     18855: gap of unknown length
18856     21177: contig of 2332 bp in length
21178     21296: gap of unknown length
21297     23327: contig of 2031 bp in length
23328     23446: gap of unknown length
23447     24500: contig of 1054 bp in length
24501     24619: gap of unknown length
24620     25921: contig of 1302 bp in length
25922     26040: gap of unknown length
```


* 26041 27519: contig of 1479 bp in length
* 27520 27638: gap of unknown length
* 27639 28757: contig of 1119 bp in length
* 28758 28876: gap of unknown length
* 28877 30055: contig of 1179 bp in length
* 30056 30174: gap of unknown length
* 30175 32305: contig of 2131 bp in length
* 32306 32424: gap of unknown length
* 32425 33869: contig of 1445 bp in length
* 33870 33988: gap of unknown length
* 33989 35107: contig of 1119 bp in length
* 35108 35226: gap of unknown length
* 35227 36536: contig of 1310 bp in length
* 36537 36655: gap of unknown length
* 36656 37886: contig of 1031 bp in length
* 37887 39387: gap of unknown length
* 39388 39506: gap of unknown length
* 39507 41435: contig of 1929 bp in length
* 41436 41554: gap of unknown length
* 41555 43232: contig of 1678 bp in length
* 43233 43351: gap of unknown length
* 43352 45009: contig of 1658 bp in length
* 45010 45128: gap of unknown length
* 45129 46195: contig of 1067 bp in length
* 46196 46314: gap of unknown length
* 46315 47965: contig of 1651 bp in length
* 47966 48084: gap of unknown length
* 48085 49583: contig of 1498 bp in length
* 49584 49701: gap of unknown length
* 49702 52099: contig of 2398 bp in length
* 52100 52218: gap of unknown length
* 52219 54677: contig of 2459 bp in length
* 54678 54796: gap of unknown length
* 54797 56319: contig of 1523 bp in length
* 56320 56438: gap of unknown length
* 56439 58355: contig of 1917 bp in length
* 58356 58474: gap of unknown length
* 58475 59651: contig of 1177 bp in length
* 59652 59770: gap of unknown length
* 59771 61723: contig of 1952 bp in length
* 61724 63273: contig of 1432 bp in length
* 63274 63392: gap of unknown length
* 63393 64657: contig of 1265 bp in length
* 64658 64776: gap of unknown length
* 64777 67435: contig of 2659 bp in length
* 67436 67554: gap of unknown length
* 67555 69652: contig of 2098 bp in length
* 69653 69771: gap of unknown length
* 69772 72205: contig of 2434 bp in length
* 72206 72324: gap of unknown length
* 72325 74056: contig of 1732 bp in length
* 74057 74175: gap of unknown length
* 74176 75928: contig of 1751 bp in length
* 75929 76045: gap of unknown length
* 76046 78421: contig of 2376 bp in length
* 78422 78540: gap of unknown length
* 78541 80884: contig of 2344 bp in length
* 80885 81003: gap of unknown length
* 81004 82691: contig of 1688 bp in length
* 82692 82810: gap of unknown length
* 82811 84563: contig of 1753 bp in length
* 84564 84682: gap of unknown length
* 84683 87125: contig of 2443 bp in length
* 87126 87244: gap of unknown length
* 87245 89423: contig of 2187 bp in length
* 89424 89530: gap of unknown length
* 89531 92882: contig of 3332 bp in length
* 92883 93001: gap of unknown length
* 93002 95515: contig of 2514 bp in length
* 95516 95634: gap of unknown length
* 95635 99067: contig of 3433 bp in length

* 99068 99186: gap of unknown length
* 99187 102713: contig of 3527 bp in length
* 102714 102832: gap of unknown length
* 102833 107003: contig of 4171 bp in length
* 107004 107122: gap of unknown length
* 107123 110602: contig of 3480 bp in length
* 110603 110721: gap of unknown length
* 110722 115096: contig of 4374 bp in length
* 115097 115214: gap of unknown length
* 115215 120208: contig of 4994 bp in length
* 120209 120327: gap of unknown length
* 120328 124050: contig of 3723 bp in length
* 124051 124170: gap of unknown length
* 124171 130483: contig of 6313 bp in length
* 130484 130602: gap of unknown length
* 130603 137345: contig of 6743 bp in length
* 137346 137464: gap of unknown length
* 144612 144730: contig of 7148 bp in length
* 144731 150942: contig of 6212 bp in length
* 150943 151061: gap of unknown length
* 151062 159108: contig of 8047 bp in length
* 159109 159237: gap of unknown length
* 159238 167672: contig of 8445 bp in length
* 167673 167791: gap of unknown length
* 167792 176185: contig of 8394 bp in length
* 176186 176304: gap of unknown length
* 176305 182851: contig of 6547 bp in length
* 182852 182970: gap of unknown length
* 182971 192307: contig of 9337 bp in length
* 192308 192426: gap of unknown length
* 192427 200688: contig of 8262 bp in length
* 200689 200807: gap of unknown length
* 200808 214576: contig of 13650 bp in length
* 214577 225057: gap of unknown length
* 225058 225176: contig of 10481 bp in length
* 225177 235058: gap of unknown length
* 235059 236928: contig of 11752 bp in length
* 236929 237047: gap of unknown length
* 237048 250963: contig of 13916 bp in length
* 250964 251083: gap of unknown length
* 251084 276109: contig of 25027 bp in length
* 276110 276227: gap of unknown length
* 276228 277416: contig of 1189 bp in length
* 277417 277534: gap of unknown length
* 277535 278530: contig of 996 bp in length
* 278531 278648: gap of unknown length
* 278649 279638: contig of 990 bp in length
* 279639 279756: gap of unknown length
* 279757 280938: contig of 1182 bp in length
* 280939 281057: gap of unknown length
* 281058 282079: contig of 1023 bp in length
* 282080 282197: gap of unknown length
* 282198 283287: contig of 1090 bp in length
* 283288 283405: gap of unknown length
* 283406 284424: contig of 1019 bp in length
* 284425 284542: gap of unknown length
* 284543 285725: contig of 1189 bp in length

alignment_scores: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-332-522b-2_COPY-335_428 x AC025242 ..

Align seg 1/1 to: AC025242 from: 1 to: 318640

33 TTTAGCGTCTTCAATGATGATGA 117846

|||||

```

seq_name: gb_sts:HSA248361
seq_documentation_block:
LOCUS      HSA248361      202 bp      DNA      STS      06-JAN-2000
DEFINITION Homo sapiens STS NL1-266R, sequence tagged site.
ACCESSION  AJ248361
VERSION    AJ248361.1  GI:5514699
KEYWORDS   STS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 202)
            Sulimova,G.E., Udina,I.G., Kunzheva,S.S. and Kompanitzhev,A.A.
            Creating NotI-STS Markers for Human Chromosome 3
JOURNAL    Mol. Biol. 33, 698-703 (1999)
REFERENCE  2 (bases 1 to 202)
            Sulimova,G.E.
            Direct Submission
            Submitted (23-JUN-1999) Sulimova G.E., Laboratory of Comparative
            Animal Genetics, N.I.Vavilov Institute of General Genetics RAS,
            Gubkin Str. 3, Moscow B-333, 117809 GSP-1, RUSSIA
            The STS markers registered were developed
            to clones from NotI library of human chromosome
            3 received by E.R. Zbarovsky (Karolinska
            Institute, Sweden) as a result of collaborative
            research work with Engelhardt Molecular Biology
            Institute of the Russian Academy of Sciences.
            The NotI clones were used to construct NotI map
            of human chromosome 3 by FISH and mapping by a
            somatic cell hybrid panel.
FEATURES
    source          location/Qualifiers
    1..202
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="3"
    /cell_type="mouse/human microcell hybrid lines"
    /cell_line="MCH903.1"
    /map="3p21.33"
    /clone_lib="NotI-linking library"
    1..21
    /note="NL1-266R forward primer"
    /PCR_conditions="96deg. 0.8min, 67deg. 0.8min, 72deg.
    0.8min"
    STS
    22..181
    /standard_name="NL1-266R"
    /note="STS PCR product"
    complement(182..202)
    /note="NL1-266R reverse primer"
    /PCR_conditions="96deg. 0.8min, 67deg. 0.8min, 72deg.
    0.8min"
BASE COUNT      38 a      62 c      73 g      26 t      3 others
ORIGIN
alignment_scores:
    Quality: 7.00      Length: 7
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-332-522B-2_COPY_335_428 x HSA248361/rev ..
Align seg 1/1 to reverse of: HSA248361 from: 1 to: 202
58 SerAlaThrLeuArgArgAla 64
|||||
92 TCGGCTACTCTCGCGCGCT 72
seq_name: gb_in1:CVU18251
seq_documentation_block:
LOCUS      CVU18251      258 bp      DNA      INV      27-JAN-1995

```

```

DEFINITION  Chortophaga viridifasciata individual 1 cytochrome b gene,
            mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION   U18251
VERSION     U18251.1  GI:624169
KEYWORDS    green-striped grasshopper.
SOURCE      Mitochondrion Chortophaga viridifasciata
ORGANISM    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
            Acridoidea; Acridoidea; Acrididae; Oedipodinae; Chortophaga.
REFERENCE   1 (bases 1 to 258)
            Chapco,W. and Martel,R.K.B.
            Mitochondrial DNA sequence variation in North American band-winged
            grasshoppers
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 258)
            Chapco,W.
            Direct Submission
            Submitted (06-DEC-1994) William Chapco, University of Regina,
            Biology, Regina, Saskatchewan, S4S 0A2, Canada
FEATURES
    source          location/Qualifiers
    1..258
    /organism="Chortophaga viridifasciata"
    /organellar="mitochondrion"
    /strain="individual 1"
    /isolate="Jameson, Saskatchewan"
    /db_xref="taxon:37250"
    /tissue_type="whole organism minus gut"
    /dev_stage="adult"
    <1..>258
    /codon_start=3
    /transl_table=5
    /product="cytochrome b"
    /protein_id="AA61672.1"
    /db_xref="GI:624170"
    /translation="IIRTLHANGSMFICILYHVGRIYGSVMYNSFTMWIGTILIF
    LVNATAMGVLPWGMSFGATGVITMLAIPDLGELVQ"
BASE COUNT      95 a      32 c      39 g      92 t
ORIGIN
alignment_scores:
    Quality: 7.00      Length: 7
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-332-522B-2_COPY_335_428 x CVU18251/rev ..
Align seg 1/1 to reverse of: CVU18251 from: 1 to: 258
31 LysLysTyrArgCysSerIle 37
|||||
44 AAAAATATAGATGCTCCATT 24
seq_name: gb_in1:CVU18252
seq_documentation_block:
LOCUS      CVU18252      258 bp      DNA      INV      27-JAN-1995
DEFINITION Chortophaga viridifasciata individual 2 cytochrome b gene,
            mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION   U18252
VERSION     U18252.1  GI:624171
KEYWORDS    green-striped grasshopper.
SOURCE      Mitochondrion Chortophaga viridifasciata
ORGANISM    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
            Acridoidea; Acridoidea; Acrididae; Oedipodinae; Chortophaga.
REFERENCE   1 (bases 1 to 258)
            Chapco,W. and Martel,R.K.B.
            Mitochondrial DNA sequence variation in North American band-winged
            grasshoppers

```

```

JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 258)
AUTHORS      Chappo, W.
TITLE        Direct Submission
JOURNAL      Submitted (06-DEC-1994) William Chappo, University of Regina,
              Biology, Regina, Saskatchewan, S4S 0A2, Canada
              Location/Qualifiers
FEATURES
  source
    1..258
      /organism="Chortophaga viridifasciata"
      /organellar="mitochondrion"
      /strain="individual 2"
      /isolate="Jameson, Saskatchewan"
      /db_xref="taxon:37250"
      /tissue_type="whole organism minus gut"
      /dev_stage="adult"
      <1..>258
        /codon_start=3
        /product="cytochrome b"
        /protein_id="AA61673.1"
        /db_xref="GI:624172"
        /translation="IIRTLHANGASMEFICILYHGRGIYGSYMTSTWMTGTTILE
        LVMATFNGYVLPWGOMSFWGAIVITNLISAIPIYGLDVLQ"

BASE COUNT   95 a          32 c          39 g          92 t

ORIGIN
seq_name: gb_in2:AF145491
seq_documentation_block:
LOCUS      AF145491      258 bp      DNA      INV      23-SEP-1999
DEFINITION Schistocerca gregaria cytochrome b gene, partial cds; mitochondrial
ACCESSION  AF145491
VERSION    AF145491.1  GI:5918551
KEYWORDS
SOURCE
  ORGANISM  Schistocerca gregaria.
            Mitochondrion Schistocerca gregaria
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
            Acridoidea; Acrididae; Acrididae; Cyrtacanthacridinae;
            Schistocerca.
            1 (bases 1 to 258)
            Chappo, W., Kuperus, W.R. and Litzenberger, G.S.
            Molecular phylogeny of melanopline grasshoppers (Orthoptera:
            Acrididae). Soc. Am. 92 (5), 617-623 (1999)
            Ann. Entomol. Soc. Am. 92 (5), 617-623 (1999)
            Chappo, W., Kuperus, W.R. and Litzenberger, G.S.
            Direct Submission
            Submitted (23-APR-1999) Biology, University of Regina, 3737 Wascana
            Parkway, Regina, Saskatchewan S4S 0A2, Canada
            Location/Qualifiers
              1..258
                /organism="Schistocerca gregaria"
                /organellar="mitochondrion"
                /db_xref="taxon:7010"
                /tissue_type="femur in alcohol"
                /note="Simon Fraser University"

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-09-332-522B-2_COPY_335_428 x CVU18252/rev ..
  Align seg 1/1 to reverse of: CVU18252 from: 1 to: 258

31 LysLysTYrAGcYsSerIle 37
|||||
44 AAAAAATATAGATGCTCCATT 24

```

```

CDS
  <1..>258
    /codon_start=3
    /transl_table=5
    /product="cytochrome b"
    /protein_id="AAD56052.1"
    /db_xref="GI:5918624"
    /translation="IIRTLHANGASMEFICILYHGRGIYGSYMTSTWMTGTTILE
    LVMATFNGYVLPWGOMSFWGAIVITNLISAIPIYGLDVLQ"

BASE COUNT   82 a          39 c          40 g          97 t

ORIGIN
seq_name: gb_sts:A0028148
seq_documentation_block:
LOCUS      A0028148      270 bp      DNA      STS      02-MAR-1999
DEFINITION Rattus norvegicus, OTSUKA clone, OT57.32/888B06, microsatellite
sequence, sequence tagged site.
ACCESSION  A0028148
VERSION    A0028148.1  GI:4518071
KEYWORDS
SOURCE
  ORGANISM  Rattus norvegicus DNA, clone:OT57.32/888B06.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
            1 (sites)
            Watanabe, T.K., Hishigaki, H., Kanemoto, N., Miyakita-Mizoguchi, A.,
            Oga, K., Okuno, S., Ono, T., Tsuji, A., Hayashi, H., Adachi, M.,
            Yamasaki, Y., Iriye, Y., Takahashi, E., Takagi, T., Nakamura, Y. and
            Tanigami, A.
            The large-scale mapping of rat microsatellite markers
            Unpublished (1998)
            2 (bases 1 to 270)
            Watanabe, T.K.
            Direct Submission
            Submitted (24-JUL-1998) to the DDBJ/EMBL/GenBank databases. Takeshi
            K Watanabe, Otsuka Pharmaceutical Co., Ltd., Otsuka GEN Research
            Institute, 463-10, Kagasuno, Kawanuchi-cho, Tokushima, Tokushima
            771-0192, Japan (E-mail:watanabe@otsuka.genome.ad.jp,
            Tel:81-886-65-2888, Fax:81-886-37-1035)
            Location/Qualifiers
              1..270
                /organism="Rattus norvegicus"
                /db_xref="taxon:10116"
                /clone="OT57.32/888B06"
                /note="OT57.32/888B06F-5'-GCCCACTGATGTGCAATACA-3',
                OT57.32/888B06R-5'-TAAAGGTGTGTGTGCAACC-3'"

BASE COUNT   61 a          37 c          73 g          96 t          3 others

ORIGIN
seq_name: gb_in2:AF145491
seq_documentation_block:
LOCUS      AF145491      258 bp      DNA      INV      23-SEP-1999
DEFINITION Schistocerca gregaria cytochrome b gene, partial cds; mitochondrial
ACCESSION  AF145491
VERSION    AF145491.1  GI:5918551
KEYWORDS
SOURCE
  ORGANISM  Schistocerca gregaria.
            Mitochondrion Schistocerca gregaria
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
            Acridoidea; Acrididae; Acrididae; Cyrtacanthacridinae;
            Schistocerca.
            1 (bases 1 to 258)
            Chappo, W., Kuperus, W.R. and Litzenberger, G.S.
            Molecular phylogeny of melanopline grasshoppers (Orthoptera:
            Acrididae). Soc. Am. 92 (5), 617-623 (1999)
            Ann. Entomol. Soc. Am. 92 (5), 617-623 (1999)
            Chappo, W., Kuperus, W.R. and Litzenberger, G.S.
            Direct Submission
            Submitted (23-APR-1999) Biology, University of Regina, 3737 Wascana
            Parkway, Regina, Saskatchewan S4S 0A2, Canada
            Location/Qualifiers
              1..258
                /organism="Schistocerca gregaria"
                /organellar="mitochondrion"
                /db_xref="taxon:7010"
                /tissue_type="femur in alcohol"
                /note="Simon Fraser University"

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-09-332-522B-2_COPY_335_428 x A0028148/rev ..

```

Align seg 1/1 to reverse of: AU028148 from: 1 to: 270

43 GlnleulysValleuLeuGys 49
|||||
186 CAGTTAAAGTGTGCGTGC 166

seq_name: gb_ro:TAUCYTB

seq_documentation_block:

LOCUS TAUCYTB 307 bp DNA ROD 23-JUN-1995
DEFINITION Sclurus tamiasclurus hudsonicus mitochondrial cytochrome b gene,
partial cds.

ACCESSION M97282
VERSION M97282.1 GI:343455

KEYWORDS cytochrome b. Mitochondrion Tamiasclurus hudsonicus (individual isolate R1)

SOURCE spleen DNA.

ORGANISM Mitochondrion Tamiasclurus hudsonicus

Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Mammalia;

Eutheria; Rodentia; Sclurognathi; Scluridae; Sclurinae;

Tamiasclurus.

1 (bases 1 to 307)

REFERENCE Wetstein,P.J., Lager,P., Jin,L., States,J., Lamb,T. and
AUTHORS Chakraborty,R.

TITLE Phylogeny of mitochondrial DNA clones in taeset-eared squirrels

JOURNAL Mol Ecol. 3 (6), 541-550 (1994)

MEDLINE 95135626

FEATURES

1..307 Location/Qualifiers

/organism="Tamiasclurus hudsonicus"

/organelle="mitochondrion"

/isolate="R1"

/db_xref="taxon:10009"

/tissue_type="spleen"

<1..>307

/codon_start=2

/transl_table=2

/product="cytochrome b"

/protein_id="AA68920.1"

/db_xref="GI:870724"

/translation="TGSLLGCLAIQIITGFLAHYTSDEMTAFSSVTHICRDVNYG
WLIRYMHANGASMEFICFLHYRGRLYGSYTFETWIGVILLFAVATAFMGVYLP

BASE COUNT 73 a 84 c 47 g 103 t

ORIGIN

alignment_scores: Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block: US-09-332-522B-2_COPY_335_428 x TAUCYTB/rev ..

Align seg 1/1 to reverse of: TAUCYTB from: 1 to: 307

ORGANISM Lama glama

Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Mammalia;

Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Lama.

REFERENCE 1 (bases 1 to 351)

AUTHORS Harmsen,M.M., Ruuls,R.C., Frenken,L.G.J. and de Geus,B.

TITLE Camelid heavy-chain VH domains consist of at least four subfamilies
with different former VL interfaces

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 351)

AUTHORS Harmsen,M.M.

TITLE Direct Submission

JOURNAL Submitted (25-JAN-1999) Harmsen M.M., Immunology Pathobiology and
Epidemiology, Institute for Animal Science and Health (ID-DLO),
Edelhertweg 15, 8200 AB Delystad, NETHERLANDS

FEATURES

1..351 Location/Qualifiers

/organism="Lama glama"

/db_xref="taxon:9844"

/rearranged

/cell_type="B-cell"

/clone="W15"

/note="isolated by RT-PCR using short hinge specific
primer lam07"

<1..>351

/gene="W15"

/note="belongs to VH3 family, subfamily I; antigen
specificity unknown"

/product="immunoglobulin VH domain"

<1..14

/gene="W15"

/note="primer VH1BACK derived"

1..351

/gene="W15"

<1..>351

/gene="W15"

/codon_start=1

/product="immunoglobulin VH domain"

/protein_id="CAB40438.1"

/db_xref="GI:4583831"

/translation="IQSGGGGLVQAGSLRLSCVAGSRTLENYRVAMFPGPKEREL
IAVINMGSDRRYVEAVQGRFNISRDNAAENTVYLLNMSLKPEDTAVYVCAAAEYGSGL
ARRDYRMGGGTQYV"

BASE COUNT 81 a 80 c 113 g 77 t

ORIGIN

alignment_scores: Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block: US-09-332-522B-2_COPY_335_428 x LGL237293 ..

Align seg 1/1 to: LGL237293 from: 1 to: 351

13 IleIleValArgArgProIys 19

|||||

272 ATTATGTGCGGCGCGGAAA 292

seq_name: gb_cm:LGL237293

seq_documentation_block:

LOCUS LGL237293 351 bp mRNA MAM 05-APR-1999

DEFINITION Lama glama rearranged immunoglobulin variable heavy chain gene

(W15), partial.

ACCESSION AJ237293

VERSION AJ237293.1 GI:4583830

KEYWORDS immunoglobulin heavy chain; immunoglobulin superfamily; variable

region.

SOURCE llama.

Thu Sep 28 07:39:52 2000

us-09-332-522b-2_copy_335_428.011.rge

Page 41

THIS PAGE BLANK (USPTO)